PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

OCT 22 1935

VOL. XXVI SEPTEMBER, 1935 NO. 6

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

The Interests of Adults 401
EDWARD L. THORNDIKE

Preferences as Indicators of Specific Academic Achievement. . . . 411
JACK W. DUNLAP

An Empirical Study of the Reliability Coefficient 416
R. C. JORDAN

Personality Traits by Fiat. II. The Consistency of the Bernreuter Personality Inventory by the Bernreuter and by the Flanagan Keys 427
IRVING LORGE, ELNA BERNHOLZ AND SAUL B. SELLS

A Comparison of Two Types of Foreign Language Vocabulary Test 435
JOHN M. STALNAKER AND WILLIAM KURATH

Emotional Factors in Reading Disabilities in School Children . . . 443
SIMON H. TULCHIN

The Out-of-school Activities of Well-adjusted and Poorly Adjusted Elementary School Pupils 455
MARTHA CRUMPTON HARDY

Who Offers What in the Psychology of Education?. 468
VEST C. MYERS

The Effect of a Group of Social Stimuli upon Attitudes. 471
IVA COX GARDNER

Publications Received 479

$6.00 per Year · Published Monthly September to May

WARWICK & YORK, INC.

BALTIMORE, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOG

Established 1910

EDITORS

JACK W. DUNLAP
Fordham University

HAROLD E. JONES
University of California

PERCIVAL M. SYMO
Teachers College,
Columbia University

H. E. BUCHHOLZ, *Managing Editor*

THE price of the Journal is $6.00 a year in the United States; $6.40 in fore countries. Part-year subscriptions are 90 cents for each number ordered. B volumes are $7.00 each; back issues are $1.10 each except when more than five y old, and then $1.20 each.

Subscribers should notify the publishers of change in address at least four we before publication of the issue with which the change is to take effect. Claims non-receipt of an issue will not be honored unless made within two weeks a receipt of the next succeeding number.

Unsolicited manuscripts should be accompanied with return postage. Ma scripts, books and other materials for review, and correspondence regarding edito and business matters should be addressed to the Publishers.

WARWICK AND YORK . *Publishers* . BALTIMORE, M

THE ACTIVITY MOVEMENT

By CLYDE HISSONG

In an attempt to overcome the weaknesses of the traditional school organization many progressive schools have developed new programs. These programs are so similar in character that collectively the changes have been referred to as the activity movement. This movement has claimed the center of the educational stage for a length of time sufficient to have engendered widespread interest in its outcomes and in its basic philosophy.

In Doctor Hissong's study an attempt has been made to discover the principles underlying the present activity movement, to determine the influence of traditional concepts in shaping the trends of the movement, and to see if in the light of the present knowledge of the child and his relation to his environment the movement rests upon a justifiable basis.

$2.00 plus 10¢ postage.

WARWICK AND YORK

Publishers BALTIMORE

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXVI | September, 1935 | Number 6

THE INTERESTS OF ADULTS[1]

I

THE PERMANENCE OF INTERESTS

EDWARD L. THORNDIKE

Institute of Educational Research, Teachers College, Columbia University

Assume that each of a group of n individuals was accurately measured in respect of the amount of his interest in each of N activities at age A_1 and similarly at age A_2. We could then answer two important questions about each interest, namely "How far does an individual's divergence from the average or modal amount of that interest in the group at age A_2 resemble his divergence from the average or modal amount of that interest in the group at age A_1?" and "How far does an individual's divergence from the average or modal amount of *his own interests* in all N activities at age A_2 resemble his divergence from the average or modal amount of *his own interests* in all N activities at age A_1?"

The two answers would both measure the permanence of interests over the interval A_1 to A_2, but in different senses. The former would tell us how far the differences among individuals in the amount or strength of any given interest are permanent characteristics of the individuals. The latter would tell us how far the pattern of interests, that is, their relative amounts within the same person, is a permanent characteristic of him.

Any cause acting to change the amount of a person's interest from A_1 to A_2 tends in and of itself to reduce the resemblance of either sort below 1.00, but under certain conditions there may be large changes

[1] The investigations reported here were made possible by a grant from the Carnegie Corporation.

with no reduction in the resemblance. If the cause in question acts upon all the n individuals in the same way, the first sort of resemblance will be unaltered. If the cause in question acts upon all the N interests of one person in the same way, the second sort of resemblance will be unaltered.

Hitherto measurements of the permanence of interests in the first sense have been hardly attempted. The lack of convenient objective measures whereby the amount of one person's liking may be compared with the amount of another person's liking for the same activity, and the energy required to secure repeated measurements over a long interval of time, have been discouraging factors. We can measure the amount or strength of an interest, as economists measure the strength of wants, by the money price a person will pay to gratify it. If a thousand persons with equal incomes were observed in respect of all their expenditures we should have valuable objective evidence. So also in the case of observations of the expenditure of time among a thousand persons who had equal amounts of it to spend. Experiments where money or time was provided with certain options for its use would be instructive. But such observations and experiments are laborious and expensive. I have not been able to make such.

I have used the easier and less trustworthy method of collecting reports of self-observations by friends and helpers who have filled out the record-sheet shown below in accordance with the instructions shown below. Such reports have the advantages of long and intimate knowledge and of simultaneous comparison of the interests concerned by the same person using the same scale. They have the disadvantage that the scale used by one person may not be the same as that used by another. The dental experiences, headaches, and chagrin which -5 signifies for A and for B may conceivably differ widely. So may the maximum enjoyment conceivable by A and by B, on the basis of which they assign ratings of $+5$. A and B may also space their -1, -2, -3, -4, $+1$, $+2$, $+3$, and $+4$ unevenly and differently. The 0 or indifference is less subject to the varying experiences of life, and may be accepted as meaning approximately the same degree of liking (namely, just not any) for all. Such reports have also the disadvantages due to individual differences in the common human tendency to think well of oneself, and to bend facts so as to maintain one's self-respect. If the likes and dislikes reported are notably creditable or discreditable some persons may deceive themselves, or the investigator, or both, much more than other persons. The very low ratings for

Item 17 (sheer idleness) may be in part due to this, and even the ratings for Items 1–16 may be somewhat distorted by it. For example, clergymen may unconsciously think they like Item 11 (my regular job) relatively better than business men do, since it is conventionally more discreditable for a clergyman to dislike saving souls than for a business man to dislike making money.

Record Sheet: Adult Interests

Degree of Interest (Liking + and Disliking −)

	20–29	30–39	40–49	50–59	60–69
1. Reading fiction	.....	.....	.....	.....	.....
2. Reading non-fiction (except newspapers)	.....	.....	.....	.....	.....
3. Reading the newspaper	.....	.....	.....	.....	.....
4. Outdoor sports, not competitive (hunting, fishing, swimming, hiking, etc.)	.....	.....	.....	.....	.....
5. Outdoor competitive games	.....	.....	.....	.....	.....
6. Sedentary games (cards, chess)	.....	.....	.....	.....	.....
7. Dancing	.....	.....	.....	.....	.....
8. Playing a musical instrument	.....	.....	.....	.....	.....
9. Listening to music	.....	.....	.....	.....	.....
10. Theatre or movies	.....	.....	.....	.....	.....
11. My regular job	.....	.....	.....	.....	.....
12. Politics	.....	.....	.....	.....	.....
13. Welfare work (in community organizations, the church, etc.)	.....	.....	.....	.....	.....
14. Talking with old friends	.....	.....	.....	.....	.....
15. Making new acquaintances	.....	.....	.....	.....	.....
16. Traveling and seeing new places	.....	.....	.....	.....	.....
17. Sheer idleness, doing nothing and thinking of nothing	.....	.....	.....	.....	.....

Instruction Sheet: Adult Interests

Please consider the activities listed on the Record Sheet from the point of view of your interest in them at the age periods noted at the top of each column. Enter a number from +5 to −5 for your degree of like or dislike for each of these activities at age twenty to twenty-nine; then do the same for age thirty to thirty-nine; then for age forty to forty-nine; then for age fifty to fifty-nine; then for age sixty to sixty-nine. You will, of course, not enter in the sixty to sixty-nine column unless you are over sixty now, and will not enter in the fifty to fifty-nine column unless you are over fifty now, etc. The meanings of the numbers from −5 to +5 are to be as follows:—

−5 Extremely unpleasant; comparable to having a dentist work on my teeth, to a bad headache, or to being made a fool of in public.

−4 Very unpleasant, but not quite so bad as −5.

−3 If you doubt whether to rate your dislike −4 or −2, call it −3.
−2 Would never do except from duty or as a means to some desired end.
−1 Mildly disliked rather than liked.
0 Indifference; neither liked nor disliked.
+1 Mildly liked.
+2 Would do without hesitation if I had the chance, and if nothing more interesting was available.
+3 If you are in doubt whether to rate your liking +2 or +4, call it +3.
+4 Very enjoyable, but not quite so much so as +5.
+5 As much liked as almost anything I can think of.

It is not expected that you will give testimony of high reliability in these ratings. If you can spare the time, please, write *A* after the numbers for which your memory is so certain that you would be willing to swear to them in a court of law, and write *X* after any numbers which you think are mere guesses. The entire record will, of course, be kept absolutely confidential.

If you would like additional copies of the Record Sheet for use with relatives or friends, I will be glad to send them and to receive the records.

We shall discuss the probable influence of these and other disturbing factors upon the apparent degree of permanence of a person's divergence from his fellow-men in the strength of the various interests, after the apparent facts have been presented, that is, the facts when the persons' reports are taken at their face value.

Taking them at their face value, and choosing Items 3, 5, 6, 8, 12, and 13 (Reading the newspaper, Competitive out-door games, Sedentary games, Playing a musical instrument, Politics, and Welfare work) for study, we find the degrees of permanence from the twenties to the fifties shown by the plots of Tables I to VI. These are for a group of one hundred twenty-two male college graduates, including forty teachers, twenty-three ministers or social workers, eight physicians, four lawyers, twenty-four psychologists, fifteen business men, and eight others (editors, engineers, etc.).* The correlations are as follows:

3 Reading newspaper	.62
5 Competitive out-door games	.69
6 Sedentary games	.71
8 Playing musical instrument	.88
12 Politics	.67
13 Welfare work	.73

The correlations for a group of sixty (twenty teachers, twenty ministers and twenty business men and lawyers, all graduates from

* Tables I and II are for only one hundred twenty-one persons.

TABLE I.—READING THE NEWSPAPER
Age twenty to twenty-nine

		−2	−1	0	1	2	3	4	5
	−2								
	−1								
Age	0	..	..	4					
50	1	..	1	..	7	2	1		
to	2	..	1	2	7	10	..		
59	3	..	..	1	10	14	15		
	4	..	..	..	6	9	3	13	
	5	..	..	..	4	..	..	5	6

TABLE II.—OUTDOOR COMPETITIVE GAMES
Age twenty to twenty-nine

		−4	−3	−2	−1	0	1	2	3	4	5
	−4	..	..	2							
	−3	..	1	1							
	−2	..	..	2	..	..	..	1			
Age	−1	..	..	..	1	1	4	..	..	..	1
50	0	..	..	..	..	11	4	3	3	4	2
to	1	..	..	1	..	..	4	2	2	4	2
59	2	..	..	..	1	..	2	5	4	11	1
	3	..	..	..	..	..	..	..	7	6	8
	4	..	..	..	..	..	..	..	3	3	5
	5	..	..	..	..	..	..	1	..	..	8

TABLE III.—SEDENTARY GAMES
Age twenty to twenty-nine

		−5	−4	−3	−2	−1	0	1	2	3	4	5
	−5	2										
	−4	1	2	1								
	−3	..	..	1	..	..	..	..	1			
	−2	..	..	..	2	..	3	..	1			
Age	−1	..	..	..	1	4	1	1	3	..	1	
50	0	..	..	..	..	1	11	4	7	1	2	
to	1	..	..	..	..	..	2	10	1	1	..	
59	2	..	..	1	1	..	3	..	8	2	2	1
	3	..	..	..	..	..	1	1	4	6	3	..
	4	..	..	1	..	..	..	2	3	3	10	..
	5	..	..	..	..	..	..	..	..	1	1	3

TABLE IV.—PLAYING A MUSICAL INSTRUMENT
Age twenty to twenty-nine

		−5	−4	−3	−2	−1	0	1	2	3	4	5
	−5	4										
	−4	..	3									
	−3	..	..	..	..	1	1					
	−2	..	..	..	3	..	..					
Age	−1	..	..	..	..	1	12	..	1			
50	0	..	..	1	..	..	44	6	2	2	3	
to	1	..	..	..	..	2	1	10	2	1	1	1
59	2	..	..	..	..	..	..	1	1	3	1	..
	3	..	..	..	..	..	..	..	..	2	1	..
	4	..	..	..	..	..	..	..	..	..	3	1
	5	..	..	..	..	..	..	..	..	..	3	4

TABLE V.—POLITICS
Age twenty to twenty-nine

		−4	−3	−2	−1	0	1	2	3	4	5
	−5	..	..	..	..	..	..	..	..	1	
	−4	2	2	1						..	
	−3	..	3	1	..	1				..	
	−2	..	1	8	2	1	1	1	1	..	
Age	−1	..	..	..	5	..	..	..	..	..	
50	0	..	..	..	1	15	2	1	..	..	1
to	1	..	..	1	..	3	12	4	1	2	..
59	2	..	..	1	1	4	..	2	1	1	..
	3	..	1	..	..	5	..	5	4	1	..
	4	..	1	..	..	..	1	1	5	9	..
	5	..	..	..	..	1	..	..	1	..	3

TABLE VI.—WELFARE WORK
Age twenty to twenty-nine

		−4	−3	−2	−1	0	1	2	3	4	5
	−4	2	..	..	..	..	..	..	1		
	−3	1	1	1	1				..		
	−2	..	..	8					..		
Age	−1	..	..	..	2	..	..	1	..		
50	0	..	..	..	..	12	1	2	..	..	1
to	1	..	..	..	..	1	8	3	3	..	1
59	2	..	..	..	..	2	3	4	7	..	..
	3	..	..	..	1	2	4	4	5	3	..
	4	..	..	..	..	3	5	2	5	8	..
	5	..	..	..	..	1	1	..	2	2	8

the same college) for these and the other items 1, 2, 4, 7, 9, etc. were as follows:

1. Reading fiction	.24
2. Reading non-fiction	.59
3. Reading the newspaper	.57
4. Outdoor sports, not competitive	.51
5. Outdoor competitive games	.61
6. Sedentary games	.76
7. Dancing	.75
8. Playing an instrument	.91
9. Listening to music	.63
10. Theatre or movies	.71
11. One's regular job	.47
12. Politics	.70
13. Welfare work	.70
14. Talking with old friends	.50
15. Making new acquaintances	.77
16. Traveling and seeing new places	.53
17. Sheer idleness	.88

If we had accurate measures of the amount of the interests instead of the ratings reported by the persons these correlations would perhaps be reduced, but not greatly.

If any person's ratings for an interest deviate from the truth by the same sort of error for the fifties as for the twenties the error will act to raise the correlation above its real value. If he rates the interest too low at one period and too high at the other, the error will act to lower the correlation. In so far as the errors of his ratings are the haphazard "chance" or "accidental" errors which afflict all human measurements, they will lower the correlation.

Differences in the scales used by different individuals will, on the whole, tend to raise an age correlation above its true value since a person's scale will tend to diverge from the truth in the same way for the fifties and for the twenties. Differences in the contamination of the ratings by the protection of one's general self-respect will also tend to raise an age correlation above its true value. Protection of one's present self-respect by a greater raising of one's ratings for desirable interests at the later than at the earlier age may, however, have the contrary effect. If a person rates himself accurately or too unfavorably for the twenties and too favorably for the fifties, his errors will be sure to reduce the age 20–50 correlation. In so far as a person for any cause assimilates his ratings for the fifties and twenties the correlations are raised above their true values. One cause might

be the belief or expectation that interests are stable and permanent; another cause might be an inertia or indolence which led a person to use the same number for the twenties as for the fifties unless there was very emphatic evidence of change; another (in the case of persons well over sixty), a general vague rating for the remote past with failure to distinguish within that past. In so far as a person for any cause other than reality differentiates his ratings for the fifties and twenties, the correlations are depressed below their true value. One such cause might be the belief or expectation that interests are shifting and transitory: another might be the desire to seem thoughtful and discriminating and a consequent unwillingness to leave many rows of identical ratings in the report, unless the evidence was emphatically to the effect that the interest was just that amount at each period. Divergences from the truth due to ordinary imperfections in observation, in memory or in inference will act as "chance" or "accidental" errors and make the age correlations lower than they should be.

The effect of these errors is probably slight except for the first and last, that is, the differences in the scales and the ordinary errors of observation, memory and inference. The first is real and probably is important; the last surely is.

We may estimate the importance of the first by observing the average ratings of individuals for Items 1–16, and allowing for the probable real differences in average liking. Suppose for example that *A* gave ratings of +5, +4, and +3, with an average rating of +4, while *B* gave ratings of +3, +2, and +1 with an average rating of +2, and that *A* was not considered to like Items 1–16 any more than *B* by persons who knew both *A* and *B*. Then probably *A* calls +5 what *B* would call +3, and calls +4 what *B* would call +2, and so on. We may also observe the variations of the ratings for Items 1–16. Suppose, for example, that *C* rated six of them as +2, four as 0, and six and −2, whereas *D* rated two of them as +5, two as +4, two as +3, one as +2, one as +1, one as −1, one as −2, two as −3, two as −4, and two as −5, and that *D* was not considered to have more pronounced likes and dislikes than *C* by persons who knew both *C* and *D*. Then probably the amounts which *C* calls +2 are called +5 or +4 or +3 (more or less by guess) by *D*, those which *C* calls 0 are scattered from +2 to −2 by *D*, those which *C* calls −2 are called −3 or −4 or −5 by *D*. If there were many cases like *A*, *B*, *C* and *D*, the influence of individual differences in the scales would obviously be very large and very important. If, on the other hand, the differences of persons

in the distributions of the ratings for Items 1–16 are largely explainable by real differences in their likes and dislikes, the influence in question is small and unimportant.

I have inspected the distributions of the ratings and compared them with my knowledge of the persons in the case of those whom I know well. The differences in the scales used seem to be moderate. For example, the two cases which show the most extreme distributions are *X* and *Y*. *X* has +5 at all four age periods (20–29, 30–39, 40–49, and 50–59) for all of the sixteen items except reading fiction, for which his average is +2, sedentary games, for which his average is 0, and dancing, for which his average is 0. His average is +4.2. *Y* averages +0.4; his ratings for the sixteen (using the average of the four periods in each case) are −2, 9, 9, 5, −12, 6, 16, 6, 13, 4, 1, 0, 0, 3, −5, 3. The difference here is largely a real difference in the persons, not in the scales they used. *X does* like very much all the activities which he rates +5. He probably would like playing cards and dancing very much, except for the fact that he is a Methodist minister. His ratings do measure him. *Y* does not get keen enjoyment out of anything except working with tools and machinery. His ratings do measure him. The evidence showing that individual differences in the scales used are moderate is presented later in the discussion of the interrelations of interests.

In my opinion the influence from the chance errors will nearly or quite balance the influence from differences in scales. So I estimate that if an omniscient and infallible observer had rated these one hundred twenty-two men, the amount of interest of a person in comparison with that of other persons would show a resemblance of about .65 between the twenties and the fifties in our group.

Our group, being limited to college graduates and containing large proportions of teachers and psychologists, has probably a smaller variance in each of these interests than the general adult male population. The correlations in the latter would, other things being equal, be higher than those in our group.

Consider now the permanence of interests in the second sense, that is, how far the relative strength of interests in the same person remains the same from the twenties to the fifties. The fact here consists of plots like those shown in Tables VII to XII. For each individual there is one such plot from which a correlation coefficient is computed. The correlations actually obtained in the case of a random sample of

Tables VII to XII.—The permanence of the *relative* strengths of sixteen interests in six individuals at or near the median of all. The scale from right to left is for age twenty to twenty-nine; the scale up and down is for age fifty to fifty-nine.

TABLE VII

	−5	−4	−3	−2	−1	0	1	2	3	4	5
−5											
−4											
−3											
−2											
−1	..	..	..	..	2						
0	..	..	..	..	..	3					
1											
2											
3	..	..	..	..	..	..	..	..	3		
4	..	..	..	..	..	..	..	..	1	2	1
5	..	..	..	..	..	..	2	1	..	1	

TABLE VIII

	−5	−4	−3	−2	−1	0	1	2	3	4	5
−5											
−4											
−3											
−2											
−1	..	..	..	..	..	..	1				
0	..	..	..	..	..	1	2				
1	..	..	..	..	..	..	2				
2	..	..	..	..	..	..	..	2			
3											
4	..	..	..	..	..	..	1	3	2	1	
5	..	..	..	..	..	..	..	..	..	..	1

TABLE IX

	−5	−4	−3	−2	−1	0	1	2	3	4	5
−5											
−4	..	1									
−3	..	..	1								
−2	..	..	..	..	..	..	1				
−1											
0	..	..	..	..	..	2					
1	..	..	..	1	1	1					
2	..	..	..	..	..	3	..	1			
3	..	..	..	..	..	1	2				
4	..	..	..	..	..	..	..	..	1		
5											

TABLE X

	−5	−4	−3	−2	−1	0	1	2	3	4	5
−5											
−4											
−3											
−2											
−1											
0	..	..	1	1	..	1					
1	..	..	..	..	..	..	..	4			
2	..	..	..	..	..	..	1	..	1		
3	..	..	..	..	..	..	..	1	1		
4	..	..	..	..	..	..	1	..	1	..	1
5	..	..	..	..	..	..	..	..	1	..	1

TABLE XI

	−5	−4	−3	−2	−1	0	1	2	3	4	5
−5											
−4											
−3											
−2	..	..	..	..	1						
−1											
0	..	..	..	..	..	1	1	1			
1	..	..	..	..	..	..	2	..	1		
2	..	..	..	..	..	1	..	4	2		
3											
4	..	..	..	..	..	..	..	..	..	1	
5	..	..	..	..	..	..	..	..	..	1	

TABLE XII

	−5	−4	−3	−2	−1	0	1	2	3	4	5
−5											
−4											
−3											
−2											
−1											
0	..	..	..	..	..	3					
1	..	..	..	..	..	..	..	..	2		
2	..	..	..	..	..	..	2	2			
3	..	..	..	..	..	..	..	1	4		
4	..	..	..	..	..	..	..	..	..	2	
5											

fifty-one persons from the graduates show a median of .75. The correlations from eighteen psychologists show a median of .74. The range is from .54 to 1.00.

Individual differences in the scales used have no influence upon these correlations. The net total influence of the various sources of error which do act upon them is, in my opinion, to reduce them below the values which would be had from infallible ratings. In my opinion the chance errors in observation, in memory and in inference, together with the errors toward unduly differentiating the twenties from the fifties will more than counterbalance the errors due to indolence or to the belief or expectation that interests will be stable. This may be doubted, but nobody who has had adequate experience with such ratings will expect that infallible measures would lower the correlations in this group by any large amount.

These correlations are also influenced by the fact that as a rule people in their fifties like reading fiction, outdoor sports, competitive outdoor games and dancing less, and like reading non-fiction, reading the newspapers, and talking to old friends more, than they did in the twenties. If we correct the ratings for the fifties to allow for this by adding 0.5, 0.5, 1.5 and 1.0 respectively to each person's ratings of 1, 4, 5, and 7 respectively, and subtracting 0.5 and 1.0 and 0.5 to each person's ratings of 2, 3, and 15 respectively, the correlations will rise.

PREFERENCES AS INDICATORS OF SPECIFIC ACADEMIC ACHIEVEMENT

JACK W. DUNLAP

Fordham University

The desirability of having individuals in situations where their talents will operate to give the maximum benefit to the individual and to society has long been recognized. The catch, however, in the problem of vocational guidance has been the difficulty of securing an exact and extensive knowledge of an individual's abilities. Certain powerful tools, such as mental tests and achievement tests, have been developed for measuring accurately some aspects of the individual. These factors, vital as they are in predicting success in any situation, lack a great deal of giving the complete picture necessary if the individual is to be directed where his efforts will have their optimum value. It is important to have a knowledge of an individual's social background, his ability to adjust to other individuals, his reaction to noises, silences, monotonous situations, confined or open work conditions, together with many other factors that act as irritants to the nervous stability of people. Finally a knowledge of the interests or preferences of each individual may be necessary. This last factor exerts a strong influence upon an individual's success in a given task or in a group of closely related tasks.

This paper concerns itself with the relationships between interests or expressed preferences of individuals and their success in academic subjects. It is necessary to start somewhere on the problem, and a logical place to measure the effect of interest would seem to be where objective, valid, and reliable measures of achievement are available. Certainly these measures are available in measuring academic achievement in the various tool subjects of the elementary school. Another reason for choosing this as a point of departure is that the results, if favorable, can be applied to the better social and academic adjustment of school children, certainly an important field of guidance. The rest of this paper is devoted to a description of a preliminary investigation and indicates the probability of the possible future use of a knowledge of interests in academic placement.

A preference blank containing four hundred thirty-five items, composed of terms from various fields of academic subject-matter was administered in May, 1933 to two groups of seventh grade pupils,

with approximately one hundred forty students in each group. Students were required to encircle one of the four symbols: *L*, *I*, *D*, or *U* for each item, thus indicating whether they liked, were indifferent to, disliked, or were unfamiliar with the term.

The terms used were selected from the various subject-matter fields on the basis of an analysis of the New York and New Jersey courses of study. If a term or phrase appeared in both courses of study the item was taken for the preliminary tryout. The items were classified by five competent judges according to the specific field to which they seemed most closely related. This classification was necessary as an item might occur in two or more subjects. For example, the term "Hawaii" might be classified from the geographical, historical, or agricultural point of view. Items such as "quotients" and "sums" were classified as arithmetic; "punctuation" and "semicolons" as grammar, etc. The items were presented in cycles composed of one term from each of the following fields: Geography, literature, arithmetic, history, general achievement, physiology, and grammar. Thus the cycle, starting with an item from geography, began on the first, eighth, fifteenth, etc. items.

Each item was then evaluated for its predictive value by comparison with the scores on the corresponding sub-tests of the New Stanford Achievement Test. Weights were determined for each of the four response categories of each item on the basis of the bi-serial correlations with the appropriate criterion. This analysis revealed that certain items had high predictive value for the specific subject-matter field from which they were drawn.

In a previous paper evidence was presented that certain items had considerable discriminative value in several fields in addition to the specific field from which they were drawn.[1] Furthermore items drawn from certain fields were found to have a much greater effective spread of prediction than do items drawn from other fields.

The next step in the study was to select from the original list those items that showed high predictive value for a given subject, for both experimental groups. These items were then combined into a new preference blank of two hundred fourteen items and administered in January, 1934, to two new groups of students to verify their effectiveness. In order to make the verification of the value of this new blank

[1] Dunlap, Jack W.: "The Predictive Value of Interest Test Items for Achievment in Various School Subjects." *Journal of Applied Psychology*, Vol. XIX, 1, 1935, pp. 53–58.

as rigorous as possible, one of these new groups consisted of one hundred forty-two eighth grade students who were measured just three days before entering the ninth grade. It will be remembered, however, that the groups on which the item weights were determined were second semester seventh grade students. The other group consisted of one hundred eighty-two first semester seventh grade students.

The criterion selected for measuring achievement was the Metropolitan Achievement Test, New York Edition, Form *A*. Thus, the items were used to predict academic success as measured by an entirely different test, for two new groups, one with a half year less training, and the other with a full year more training than the original experimental groups. The time necessary to administer the preference blank is approximately twenty minutes, or on the average three minutes for each school subject.

The Terman Group Test of Mental Ability was also administered to these groups.

The preference blanks were then scored for arithmetic, geography, grammar, history, literature and general achievement. The blank was not scored for physiology as the Metropolitan test did not include this subject. The correlations between the preference scores for each academic subject and the corresponding sub-test of the Metropolitan test was determined for each grade group. These correlations are shown in Table I.

The average correlation between the various achievement scores and preference scores is .47 for the seventh grade and .42 for the eighth grade. The correlations for geography, grammar, history, and general achievement are substantial, note column one for both the seventh and eighth grade groups. The smallest value, .40, occurs for geography for the eighth grade group and the highest value, .60, for seventh grade general achievement.

The correlations for literature fluctuate more widely than for any of the other subjects. The correlations for arithmetic would indicate that there is less promise of developing a satisfactory preference blank for predicting arithmetic than for any of the other subjects.

In column two the correlations between the achievement scores and the total score on the Terman Group Test of Mental Ability are given. In every instance the Terman Group Test correlates higher with the sub-tests of the achievement test than do the scores on the preference blank. This is not unexpected as it is well known that the scores on group intelligence and achievement tests have high intercor-

TABLE I.—THE INTERCORRELATIONS BETWEEN (1) CERTAIN SUB-TESTS OF THE METROPOLITAN ACHIEVEMENT TEST, (2) EXPRESSED PREFERENCES AND (3) THE TERMAN GROUP TEST OF MENTAL ABILITY. MULTIPLE AND PARTIAL CORRELATIONS ARE GIVEN IN COLUMNS FOUR AND FIVE

Subject	Grade	r_{12}	r_{13}	r_{23}	$r_{1.23}$	$r_{12.3}$
Literature	7	.56	.63	.45	.70	.40
	8	.35	.54	.37	.56	.19
Arithmetic	7	.31	.54	.28	.57	.20
	8	.30	.46	.38	.48	.15
Geography	7	.46	.58	.30	.65	.37
	8	.40	.62	.36	.65	.24
Grammar	7	.48	.69	.54	.70	.18
	8	.45	.72	.46	.73	.19
History	7	.43	.68	.25	.73	.37
	8	.48	.67	.41	.71	.30
General achievement	7	.60	.85	.53	.87	.34
	8	.53	.81	.55	.82	.17

relations and what is more important here, the score on the preference blank for any particular subject represents only three minutes of testing time, whereas the score on the Terman test represents approximately thirty minutes. This difference of time in favor of the preference blank puts such blanks in a much more favorable light than would appear at a casual inspection of the correlation coefficient. The correlations of .81 and .85 between the total scores on the achievement test and the intelligence test indicate a high degree of community of function between the Terman and the Metropolitan tests.

Column three presents the correlations between the scores of the sub-tests of the preference blank with the score on the Terman test.

In columns four and five, respectively, the multiple and partial correlations are presented. In each case the increase in the multiple correlation has been in the second decimal place, and although quite small in certain instances, notably grammar, it is still appreciable when it is considered that this increase was secured by one three minutes additional testing time. The average increase for the seventh grade is .04 with a range of .01 to .07.

The increase was least for grammar and greatest for geography and literature. The average increase for the seventh grade was .02, with a range from .01 to .04. The increase in the multiple correlation was four points for history but only one point for grammar and general achievement.

The high degree of relationship between the achievement and intelligence tests as shown by the correlations in column two tend to mask the predictive value of the preference measures. The value of the preference scores is shown more clearly in column five where the partial correlations between achievement and expressed preferences, independent of the effects of intelligence, are given. These partial correlations are all positive. The expectancy that the smallest of these, .15 for eighth grade arithmetic, would become zero or negative in another sample is one in thirty.

This report may be summarized as follows: Two courses of study were analyzed and basic items occurring in both were selected from several subject-matter fields. These items were arranged in a preference blank and administered to two groups of students together with criterion tests. After a statistical analysis certain items were selected for further study and were administered together with an achievement test and an intelligence test to two new groups of students. Zero order correlations, partial correlations, and multiple correlations were determined for the twelve sets of data. These correlations were all positive. In conclusion it would seem that if the preliminary form of the preference blank described above were extended and refined, the expressed preferences of an individual could be used to increase materially the accuracy of the prediction of future academic success at the junior high school level.

AN EMPIRICAL STUDY OF THE RELIABILITY COEFFICIENT

R. C. JORDAN

Wolvin Elementary School, Texas City, Texas

I. INTRODUCTION

Purpose.—This paper is devoted to a study of differences in coefficients of test reliability with different methods of computation. The problem arose during the process of constructing and evaluating a test for scholastic aptitude. As will be explained later, results from the administration of this test supply the numerical data for the study.

Related Material.—It has long been recognized that actual measures in psychology and education differ more or less from objective reality. In 1904, Spearman,[1] presenting in an article his formula for correction for attenuation in finding the relation between two variables, writes:

> Now, suppose that we wish to ascertain the correspondence between a series of values, p and another series, q. By practical observation we do not obtain the true objective values p and q, but only approximations which we will call p' and q'. Obviously p' is less closely connected with q' than is p with q, for the first pair only correspond at all by the intermediation of the second pair. The real correspondence between p and q, shortly r_{pq} has been "attenuated" into $r_{p'q'}$.

In his next paragraph is an expression of the mathematical relationship which later came into common use as a measure of reliability of a test. He states:

> To ascertain the amount of this attenuation and thereby discover the true correlation, it appears necessary to make two or more independent series of observations both of p and q. Then,

$$r_{pq} = \frac{r_{p'q'}}{(r_{p'p'} \times r_{q'q'})^{\frac{1}{2}}}$$

> Where $r_{p'q'}$ = the mean of the correlations between each series of values obtained for p with each series obtained for q.
> $r_{p'p'}$ = the average correlation between one and another of these several independently obtained series of values for p.
> $r_{q'q'}$ = the same as regards q.
> and r_{pq} = the required real correlation between the true objective values of p and q.

---

[1] Spearman, C.: "The Proof and Measurement of Association Between Two Things." *American Journal of Psychology*, Vol. XV, 1904, p. 90.

The $r_{p'p'}$ and $r_{q'q'}$ are really formulae for measuring the reliability of a test. Whipple,[1] in 1914, writes:

> Attention should be called to the opportunity offered by such correlations as A_1A_2 and B_1B_2 to secure an objective indication of the reliability of the tests used to measure A and B. The correlation thus measured between results of two different applications of the same test upon the same persons has been used, particularly by recent investigators like Spearman, Burt, Wyatt, *et al.*, as the coefficient of reliability. The principle is simple enough. If the outcome of the test is not disturbed by chance or by constant errors, the rank of the several S's should be the same at each trial.

Whipple states also the possibility of using the odd-even combinations in constituting the variables to be correlated.

The present concept of test reliability is represented by the definition given by Otis,[2] who says:

> The term reliability is used technically in connection with tests to mean the degree to which a test is consistent in measuring that which it measures.

Other authors of statistical textbooks define test reliability in essentially the same manner.

Garrett[3] lists three ways of finding the reliability coefficient.

1. Correlate the scores made on the test with the scores made on the same test by the same group.
2. Correlate the scores made on the test with the scores made on a duplicate test by the same group.
3. Correlate halves of the test and find the reliability of the whole test by application of the Spearman-Brown formula.

So far as it is indicated in the discussion it may be assumed that the three techniques are regarded as equivalent, that it, that approximately the same reliability coefficient will result from the different techniques.

Two recent articles, however, show a growing appreciation of the possibility of differences in reliability coefficients resulting from different methods of derivation. In the first of these, Remmers[4] discusses the three techniques and says that the duplicate-form method and the split-halves method are most commonly used. He says further:

---

[1] Whipple, G. M.: *Manual of Mental and Physical Tests, Part I Simpler Processes*, pp. 52–53.

[2] Otis, A. S.: *Statistical Method in Educational Measurement*, p. 227.

[3] Garrett, H. E.: *Statistics in Psychology and Education*, pp. 268–271.

[4] Remmers, H. H.: "A Possible Experimental Error in Determining the Overlap of Two Correlated Variables." *Journal of General Psychology*, Vol. IX, 1933, pp. 459–461.

It is of crucial importance to note, however, that these two methods do not, in general, yield the same reliability coefficient. The reliability obtained by the split-test method will in general be higher than that obtained from the equivalent-forms method.

He states also:

Such factors as fatigue, monotony, distraction, etc., will be more likely to affect reliability in the latter [equivalent-forms] technique . . .

Dunlap,[1] in discussing these three methods of finding reliability, says that the reliability coefficient secured by giving equivalent forms and then correlating the scores measures the variability of the subjects as well as the reliability of the test. He believes that the method of correlating the halves of a test gives the truer reliability. He says:

The Spearman-Brown formula gives the reliability of the test independent of the reliability of the subjects.

He says further:

. . . the true score on the Spearman-Brown formula is the "true" ability at the instant, while in the "true" score of the intercorrelation of two forms is the true underlying ability of the subject.

Again he states:

The use of the correlation between forms of a test as the measure of reliability of one of them involves more assumptions than the Spearman-Brown formula for determination of the reliability of a single form.

Dunlap does not, however, bring any definite data to the support of these conclusions.

Problem.—In the study here reported data have been collected to give an empirical answer to some of the questions which arise when different methods are used in deriving coefficients of reliability. Specifically, the problem has been to answer the following questions:

1. Is there any difference between coefficients of reliability obtained by correlating duplicate forms of a test and those obtained by correlating alternate items in a single form?

2. Is there any difference between reliability coefficients determined by correlating even-odd items and using the Spearman-Brown formula and coefficients determined by actually doubling the length of the test?

3. If differences are found, which coefficient probably most nearly represents the true reliability of a test?

[1] Dunlap, J. W.: "Comparable Tests and Reliability." *Journal of Educational Psychology*, Vol. XXIV, 1933, pp. 442–453.

II. PROCEDURE

The instrument used in this study was the Scholastic Aptitude Test. This test was prepared in two equivalent forms, *A* and *B*, by H. T. Manuel, James Knight, Woodie Mae Anderson, Edna Mae Boles, J. A. Floyd and Roland C. Jordan. Each form consists of three parts: Part I, Word Meaning; Part II, Word Relations; Part III, Mathematics. The test is self administering and required ten minutes for each part or a total time of thirty minutes to administer. It is suitable for use at the high school and college freshman levels. Table I, below, gives the means and standard deviations of the two forms at the college freshman and high school senior level.

TABLE I.—MEANS AND STANDARD DEVIATIONS OF SCHOLASTIC APTITUDE TEST

Level	Form	No.	Means	Standard deviations
College freshman	*A*	210	56.66	15.82
	B	210	56.81	14.65
High school senior	*A*	200	45.06	12.07
	B	200	44.37	11.84

Both forms of the test were given to several hundred college freshmen and high school seniors. They were administered in *ABBA* order, that is, one half of the group was given Form *A* first followed by Form *B* and the other half was given Form *B* first followed by Form *A*. Both forms of the test were given at one sitting. For the purpose of this study, one hundred five papers were selected by chance from among the college freshmen taking the test in *AB* order. The same number of papers were selected from the college freshmen taking the test in *BA* order. At the high school senior level, one hundred papers were selected from the group taking the test in *AB* order and the same number taking it in *BA* order.

These tests were scored in several ways. First, by forms; then each form was scored for even-odd items; then the even and odd items of both forms were combined into one even-item score and one odd-item score. Consequently, the following scores were secured for each subject:

(*a*) Score on Form *A*.
(*b*) Score on Form *B*.
(*c*) Score on the even items of Form *A*.

(*d*) Score on the odd items of Form *A*.
(*e*) Score on the even items of Form *B*.
(*f*) Score on the odd items of Form *B*.
(*g*) Score on even items of Forms *A* and *B* together.
(*h*) Score on odd items of Forms *A* and *B* together.

From these scores ten reliability coefficients were calculated. In Table II, below, these coefficients are described. Letters have been used to designate each coefficient so that references to them can be more easily made in the discussion.

TABLE II.—LIST OF RELIABILITY COEFFICIENTS CALCULATED FROM SCORES ON THE SCHOLASTIC APTITUDE TEST ADMINISTERED TO TWO HUNDRED TEN COLLEGE FRESHMEN IN ABBA ORDER

Coefficient	Description	Cases
Q	Form *A* with Form *B* of tests given in *AB* order	105
R	Form *A* with Form *B* of tests given in *BA* order	105
S	Form *A* with Form *B* of tests given in *ABBA* order	210
T	Even items with odd items of Form *A* given in *BA* order	105
U	Even items with odd items of Form *B* given in *BA* order	105
V	All even items with all odd items in Forms *A* and *B* given in *BA* order	105
W	Even items with odd items of Form *A* given in *AB* order	105
X	Even items with odd items of Form *B* given in *AB* order	105
Y	All even items with all odd items in Forms *A* and *B* given in *BA* order	105
Z	All even items with all odd items in Forms *A* and *B* given in *ABBA* order	210

The same coefficients were calculated from the scores on tests given to high school seniors. Letter designations are the same except that in the latter case they are written with a prime so as to distinguish them from those used to designate the coefficients of the college freshmen.

The coefficients were determined by using the product-moment formula. In the case of coefficients based upon the split half of a single form in which only half as many items were available, the coefficients were "stepped up" by the use of the Spearman-Brown formula:

$$r_x = \frac{Nr}{1 - (N - 1)r}$$

The use of this formula makes all the coefficients comparable as regards length of test. For those correlations stepped up with the Spearman-Brown formula, the following formula suggested by Shen[1] was used.

$$SD_{r11} = 2(1 - r_{11})/N^{1/2}$$

Two formulae were used in calculating the standard error of the difference between two correlation coefficients. To find the standard error between two uncorrelated coefficients the following formula was used.

$$SD_{(r_1 - r_2)} = (SD^2_{r_1} + SD^2_{r_2})^{1/2}$$

For those coefficients that were correlated the formula shown below was used to calculate the coefficient of correlation.

$$r_{r_{12}r_{34}} = \frac{(r_{13} - r_{12}r_{23})(r_{24} - r_{23}r_{34}) + (r_{14} - r_{13}r_{34})(r_{23} - r_{12}r_{13}) + (r_{13} - r_{14}r_{34})(r_{24} - r_{12}r_{14}) + (r_{14} - r_{12}r_{24})(r_{23} - r_{24}r_{34})}{2(1 - r^2_{12})(1 - r^2_{34})}$$

After getting the correlation between the correlation coefficients, the following formula was then applied to secure the standard error of the difference.

$$SD_{(r_{12} - r_{34})} = (SD^2_{r_{12}} + SD^2_{r_{13}} - 2r_{r_{12}r_{13}}SD_{r_{12}}SD_{r_{34}})^{1/2}$$

As the measure of the reliability of a difference between two coefficients, the ratio of that difference to the standard error of the difference was taken. Any such measure is, of course, relative. One difference can not be said to be reliable and another difference unreliable. All that can be said is that one difference is *more* reliable than another. However, for the purpose of this study, the critical ratio has been taken at three. In other words, if the ratio of the difference between two coefficients to the standard error of that difference is three or more, the difference is said to be reliable. This is in agreement with general practice.

III. RESULTS

The ten reliability coefficients and their standard errors calculated according to methods previously described are shown in Tables III and IV just above the double line. Below this line are the differences between each two of these coefficients, the standard error of this

[1] Shen, E.: "The Standard Error of Certain Estimated Coefficients of Correlation." *Journal of Educational Psychology*, Vol. XV, 1924, pp. 462–465.

TABLE III.—SUMMARY OF COMPARISON OF TEN RELIABILITY COEFFICIENTS DERIVED FROM TESTS GIVEN COLLEGE FRESHMEN

	Correlation of forms			Correlation of odd items with even items						
Designation	*Q*	*R*	*S*	*T*	*U*	*V*	*W*	*X*	*Y*	*Z*
Form used	*A & B*	*A & B*	*A & B*	*A*	*B*	*A & B*	*A*	*B*	*A & B*	*A & B*
Order given	*AB*	*BA*	*ABBA*	*BA*	*BA*	*BA*	*AB*	*AB*	*AB*	*ABBA*
No. of cases	105	105	210	105	105	105	105	105	105	210
Correlation	.91	.87	.86	.94	.94	.94	.93	.88	.92	.93
SD_r	.016	.023	.018	.012	.012	.011	.014	.023	.014	.009
Q D_r	....	.04	.05	.03	.03	.03	.02	.03	.01	.02
SD_D	....	.028	.024	.019	.019	.019	.015	.020	.015	.018
$Rel._D$	....	1.42	2.08	1.58	1.58	1.58	1.33	1.50	0.67	1.11
R D_r	....	....	.01	.07	.07	.07	.06	.01	.05	.06
SD_D	....	....	.029	.021	.021	.020	.026	.031	.027	.024
$Rel._D$	....	....	0.34	3.33	3.33	3.50	2.31	0.32	1.85	2.50
S D_r	....	....	....	.08	.08	.08	.07	.02	.06	.07
SD_D	....	....	....	.021	.021	.021	.022	.027	.022	.015
$Rel._D$	....	....	....	3.81	3.81	3.81	3.18	0.74	2.72	4.67
T D_r	....	....	....	....	.00	.00	.01	.06	.02	.01
SD_D	....	....	....	....	.015	.011	.017	.023	.017	.014
$Rel._D$	....	....	....	....	0.00	0.00	0.58	2.60	1.17	0.71
U D_r	....	....	....	....	....	.00	.01	.06	.02	.01
SD_D	....	....	....	....	....	.011	.017	.023	.017	.014
$Rel._D$	....	....	....	....	....	0.00	0.58	2.60	1.17	0.71
V D_r	....	....	....	....	....	....	.01	.06	.02	.01
SD_D	....	....	....	....	....	....	.017	.023	.017	.014
$Rel._D$	....	....	....	....	....	....	0.58	2.60	1.17	0.71
W D_r	....	....	....	....	....	....	....	.05	.01	.00
SD_D	....	....	....	....	....	....	....	.024	.013	.015
$Rel._D$	....	....	....	....	....	....	....	2.08	0.77	0.00
X D_r	....	....	....	....	....	....	....	....	.04	.05
SD_D	....	....	....	....	....	....	....	....	.017	.022
$Rel._D$	....	....	....	....	....	....	....	....	2.35	2.27
Y D_r	....	....	....	....	....	....	....	....	....	.01
SD_D	....	....	....	....	....	....	....	....	....	.016
$Rel._D$	....	....	....	....	....	....	....	....	....	0.62

TABLE IV.—SUMMARY OF COMPARISON OF TEN RELIABILITY COEFFICIENTS DERIVED FROM TESTS GIVEN HIGH SCHOOL SENIORS

	Correlation of forms			Correlation of odd items with even items						
Designation	*Q'*	*R'*	*S'*	*T'*	*U'*	*V'*	*W'*	*X'*	*Y'*	*Z'*
Form used	*A & B*	*A & B*	*A & B*	*A*	*B*	*A & B*	*A*	*B*	*A & B*	*A & B*
Order given	*AB*	*BA*	*ABBA*	*BA*	*BA*	*BA*	*AB*	*AB*	*AB*	*ABBA*
No. of cases	100	100	200	100	100	100	100	100	100	200
Correlation	.81	.85	.80	.91	.94	.93	.94	.89	.89	.95
SD_r	.034	.027	.025	.018	.012	.013	.012	.022	.021	.077
Q' D_r	....	.04	.01	.10	.13	.12	.13	.08	.08	.14
SD_D	....	.043	.042	.038	.035	.036	.027	.032	.033	.034
$Rel._D$	....	0.93	0.24	2.63	3.71	3.33	4.81	2.50	2.42	4.11
R' D_r	....	....	.05	.06	.09	.08	.09	.04	.04	.10
SD_D	....	....	.036	.024	.023	.023	.029	.034	.034	.027
$Rel._D$	....	....	1.39	2.50	3.91	3.48	3.10	1.17	1.17	3.70
S' D_r	....	....	....	.11	.14	.13	.14	.09	.09	.15
SD_D	....	....	....	.030	.027	.028	.027	.032	.032	.020
$Rel._D$	....	....	....	3.67	5.18	4.64	5.18	2.81	2.81	7.50
T' D_r	....	....	....	....	.03	.02	.03	.02	.02	.04
SD_D	....	....	....	....	.020	.016	.020	.027	.027	.018
$Rel._D$	....	....	....	....	1.50	1.25	1.50	0.74	0.74	2.22
U' D_r	....	....	....	....	....	.01	.00	.05	.05	.01
SD_D	....	....	....	....	....	.014	.015	.024	.024	.013
$Rel._D$	....	....	....	....	....	0.71	0.00	2.08	2.08	0.77
V' D_r	....	....	....	....	....	....	.01	.04	.04	.02
SD_D	....	....	....	....	....	....	.017	.025	.025	.015
$Rel._D$	....	....	....	....	....	....	0.59	1.60	1.60	1.33
W' D_r	....	....	....	....	....	....	....	.05	.05	.01
SD_D	....	....	....	....	....	....	....	.024	.019	.013
$Rel._D$	....	....	....	....	....	....	....	2.08	2.63	0.77
X' D_r	....	....	....	....	....	....	....	....	.00	.06
SD_D	....	....	....	....	....	....	....	....	.022	.022
$Rel._D$	....	....	....	....	....	....	....	....	0.00	2.72
Y' D_r	....	....	....	....	....	....	....	....	....	.06
SD_D	....	....	....	....	....	....	....	....	....	.022
$Rel._D$	....	....	....	....	....	....	....	....	....	2.72

difference and the ratio of each difference to its standard error. These values are found at the intersection of the lines with the columns. To compare any two coefficients, find the one coming first alphabetically in the column on the left; follow that line until the column headed by the other coefficient is reached. In the small rectangle at this intersection are three values. The first is the difference between the two coefficients being compared. The second is the standard error of that difference and the third is the reliability of the difference. Suppose coefficients *S* and *W* are to be compared. *S* is found in the column on the left. The "*S*" line is followed until the "*W*" column is reached. There it is found that the difference between coefficients *S* and *W* is .07; the standard error of the difference is .022 and the ratio of the difference to the standard error of the difference, or the reliability of the difference, is 3.18.

The first comparison to be made is that between coefficients secured by correlating forms and those secured by correlating split halves. Comparisons are made between comparable scores, that is, between scores given to the same groups in the same order. This comparison is made in Table V below.

TABLE V.—DIFFERENCE BETWEEN COEFFICIENTS DERIVED FROM CORRELATING SPLIT-HALVES AND COEFFICIENTS DERIVED FROM CORRELATING EQUIVALENT FORMS

College Freshman Level	High School Senior Level
$T - R = \underline{.07}$	$T' - R' = .06$
$U - R = \underline{.07}$	$U' - R' = \underline{.09}$
$V - R = \underline{.07}$	$V' - R' = \underline{.08}$
$W - Q = .02$	$W' - Q' = \underline{.13}$
$X - Q = .03$	$X' - Q' = .08$
$Y - Q = .01$	$Y' - Q' = .08$
$Z - S = \underline{.07}$	$Z' - S' = \underline{.15}$

In this table, those differences that are reliable according to the accepted criterion are underlined. It is seen in the first place that in *every* case the coefficient derived from the split-halves technique is higher than the comparable coefficient derived from the equivalent-forms technique. There are fourteen comparisons made. In eight of these the difference is reliable. An examination of Tables III and IV shows, furthermore, that there are no reliable differences between any two coefficients derived from correlating split-halves. These facts, combined with the fact that the differences are all in the

same direction, answer with considerable assurance the first question around which this study was centered.

It is evident that there is a difference between coefficients of reliability derived from the correlation of equivalent forms and those derived from the correlation of split halves. It is also evident that this difference is in favor of the split halves technique. It can easily be shown why this is true. The differences in test results on two different administrations are due to several factors. Among these are the reliability of the test and the variability of the students. The variability of the student is due to attitudinal, emotional, and physical factors. When two forms of a test are given, this variability on the part of the student is apparent. This is true even when one form follows the other directly. Some students become bored; some become resentful; some are motivated to try harder; some do not try as hard. The fact remains that correlations between two forms of a test measures the variability of the student as well as the reliability of the test. When one form of a test is given and reliability derived from correlating alternate items, it can be assumed that the variability of the student has been eliminated. It is very improbable that a student's attitude should change from item to item in such a way that even items will be affected differently from odd items. It is reasonable, then, to assume that the coefficient derived from correlating odd-even items of a test most nearly represents the true reliability of the testing instrument. The objection might be raised that the measure of pupil variability is one of the things desired. However, if such a measure is desired and is secured, it should be designated as a measure of pupil variability and not as a measure of test reliability.

The second comparison to be made is that between "stepped-up" coefficients obtained by correlating even-odd items in a single form and coefficients derived by combining the even-odd items of both forms. This is in truth a measure of the validity of the Spearman-Brown formula. Table VI shows this comparison.

TABLE VI.—DIFFERENCE BETWEEN STEPPED-UP COEFFICIENTS OF EVEN-ODD CORRELATIONS OF SINGLE FORM AND COMBINED EVEN-ODD ITEMS OF BOTH FORMS

College Freshman Level	High School Senior Level
$Y - W = -.01$	$Y' - W' = -.05$
$Y - X = .04$	$Y' - X' = 0$
$V - T = 0$	$V' - T' = .02$
$V - U = 0$	$V' - U' = -.01$

It is evident from Table VI that the coefficients derived from the application of the Spearman-Brown formula do not differ reliably from the coefficients derived from actually doubling the length of test. There are eight comparisons made. Three of the differences are zero. Three of the remaining differences are negative and two are positive. None of the differences is statistically reliable. This comparison answers the second question mentioned in the discussion of the problem.

As to the third question around which this study was centered, it can be said with some assurance that the coefficients derived from correlating the split halves of a test more nearly represents the true reliability that the coefficients derived from correlating duplicate forms. The basis for this statement has been discussed more fully in the answer to the first question. It is principally due to the fact that in the split-halves technique, the pupil variability has been eliminated.

IV. CONCLUSIONS

In so far as generalizations can be made from a study of this type, the results obtained justify the following conclusions:

1. The coefficient of reliability secured from correlating odd items with even items is higher than that secured by the duplicate-forms technique.
2. The coefficient derived from correlating even items with odd items is probably a truer measure of the reliability of a test than that obtained from correlating duplicate forms because the even-odd method eliminates pupil variability.
3. The coefficient of reliability secured by applying the Spearman-Brown formula to the coefficients derived from correlating split halves is essentially the same as that derived from actually doubling the length of the test.
4. Textbooks in statistics and measurement should differentiate between the different techniques of securing the reliability of a test.
5. The meaning of a reliability coefficient lacks precision unless the method used in determining it is described.

PERSONALITY TRAITS BY FIAT. II. THE CONSISTENCY OF THE BERNREUTER PERSONALITY INVENTORY BY THE BERNREUTER AND BY THE FLANAGAN KEYS[1]

IRVING LORGE

Institute of Educational Research, Teachers College, Columbia University

WITH THE ASSISTANCE OF

ELNA BERNHOLZ AND SAUL B. SELLS

Columbia University

The Bernreuter Personality Inventory has been subjected to an analysis in which it was demonstrated that the intercorrelations among the four total trait scores were of the same order of magnitude as the intercorrelations among the trait keys. The intercorrelations among the trait keys were computed separately for "Yes" weights, for "No" weights, and for "?" weights.[2] Stagner[3] has shown that the reliability of the separate trait scores is .88 for B1-N, neurotic tendency; .80 for B2-S, self sufficiency; .87 for B3-I, introversion-extroversion; and .85 for B4-D, dominance-submission. These two studies indicate first, that the traits as measured by the Bernreuter Personality Inventory are not independent, or pure, or specific; second, that the community of traits is in the keys, and hence, may not be in the subjects; and third, that the reliability is good enough for group differentiation, but of too low an order for individual diagnosis or prognosis. A test, before its validity is to be demonstrated must show, at least, reliability of a specific magnitude consonant with the task the test is

[1] Acknowledgment is hereby made of the services rendered by the personnel furnished by the Works Division, Emergency Relief Bureau of New York City on Project 89FB-125X.

This study is part of a larger study in Interests, Attitudes and Motives supported in part by a grant from the Columbia University Council for Research in Social Science.

[2] Lorge, Irving: "Personality traits by fiat. I. The analysis of the total trait scores and keys of the Bernreuter Personality Inventory." *Journal of Educational Psychology*, Vol. XXVI, 1935, pp. 273–278.

[3] Stagner, Ross: "Validity and the reliability of the Bernreuter Personality Inventory." *Journal of Abnormal and Social Psychology*, Vol. XXIII, 1934, pp. 413–418.

to do; consistency within its modes of response; and, in the case of a multi-trait instrument, independence of the traits within the keys.

It is the purpose of this study to examine the consistency of the Bernreuter Personality Inventory. The Bernreuter Personality Inventory allows the subject three modes of response to each question: "Yes," "No," and "?." The keys are arranged in such a manner that each response should contribute to the person's trait score. From such a scoring system, it is to be expected that all the "Yes" responses will contribute positively to the total trait score, as should the "No" responses and the "?" responses. It is, further, to be expected that the "Yes" responses and the "No" responses and "?" responses, since they are measuring the same specific trait, should be correlated among themselves.[1] If neurotic tendency is a trait, the average of the weights for "Yes" responses as measured by B1-N should be correlated with the average of the weights for "No" responses as measured by B1-N and with the average of the weights for "?" responses as measured by B1-N.

The analysis to which the Bernreuter Personality Inventory scores was subjected involved the computation of twelve separate scores for each individual. A score was made up for each individual of the average of the weights earned in any trait by the "Yes" responses; another score was made up for each individual of the average of the weights earned in any trait by the "No" response; and still another score was made up for each individual of the average of the weights earned in any trait by the "?" responses. Thus, for each individual there would be twelve scores: a "Yes," a "No" and a "?" average trait-response score for each of the four traits presumed to be measured

---

[1] Willoughby: ("The concept of reliability." *Psychological Review*, Vol. XLII, 1935, pp. 153–165) has been concerned with the same problem. He writes "Concomitant variation between two items must, it seems to us, be presumptive evidence of allied causation. It would be very difficult indeed to imagine a state of affairs in which two variables were genuinely correlated in the absence of all causal connection. The causal pathways, roughly, may be of three sorts: *A* may cause *B*, *B* may cause *A*, or a third factor *C* may cause both. Since in the usual test situation we hypothesize an 'ability' and endeavor to formulate items which will 'test' or manifest that ability, the most reasonable assumption of the three in the case of the intercorrelation of test items is that it arises because both responses are determined by the ability in question. It seems to us that such a proposition will warrant conversion: The presence of high intercorrelation between two items may be taken as evidence that both are representative manifestations of a common ability or trait—a unitary trait, if one wishes."

by the Bernreuter Personality Inventory, *i.e.*, B1-N, neurotic tendency; B2-S, self-sufficiency; B3-I, introversion-extroversion; and B4-D, dominance-submission.

The population used in this study is the same as reported by Lorge previously. The Bernreuter Personality Inventory had been administered to one hundred three adults regularly employed as subjects in an adult learning project. These one hundred three subjects were adults of both sexes ranging in age from twenty to seventy, and ranging in intellect from CAVD three hundred sixty-nine to CAVD four hundred forty (on Army Alpha, the range was from thirty-eight to two hundred seven).

The sixty-six intercorrelations among the twelve trait responses are reported in Table I. The first intercorrelations to be inspected will be those between the average trait-response scores for "Yes" and for "No" responses in each trait. The correlations are +.69, +.25, +.47 and +.66 for B1-N, B2-S, B3-I, and B4-D, respectively. From

TABLE I.—THE INTERCORRELATIONS OF THE AVERAGE TRAIT-RESPONSE SCORES BY THE BERNREUTER KEYS FOR THE BERNREUTER PERSONALITY INVENTORY: FOR ONE HUNDRED THREE EDUCATED ADULTS

where

N-Yes = average trait-response score computed for each individual's "Yes" responses as measured by the B-1-N Key.

N-No = average trait-response score computed for each individual's "No" responses as measured by the B-1-N Key.

N-? = average trait-response score computed for each individual's "?" responses as measured by the B-1-N Key.

and where

S-Yes, S-No, and S-? have similar meanings as measured by the B-2-S Key.

I-Yes, I-No, and I-? have similar meanings as measured by the B-3-I Key.

D-Yes, D-No, and D-? have similar meanings as measured by the B-4-D Key.

	N-Yes	N-No	N-?	S-Yes	S-No	S-?	I-Yes	I-No	I-?	D-Yes	D-No	D-?
N-Yes	..	.25	.20	−.32	−.25	−.06	.72	.68	.07	−.78	−.50	−.05
N-No	..	...	.18	−.28	−.19	−.07	.48	.55	.04	−.56	−.45	−.04
N-?	..	...	...	.01	−.01	.11	.17	.22	.57	−.27	−.20	.34
S-Yes	..	...	...	.....	.69	.07	−.30	−.22	.08	.53	.22	−.05
S-No	..	...	...	.....	.....	.18	−.16	−.11	.11	.40	.32	.06
S-?	..	...	...	.....	.....	.....	−.07	−.05	−.00	.04	.07	.44
I-Yes	..	...	...	.....	.....	.....	.....	.47	.12	−.65	−.37	−.06
I-No	..	...	...	.....	.....	.....	.....	.....	.15	−.70	−.56	−.05
I-?	..	...	...	.....	.....	.....	.....	.....	.....	−.11	−.12	.30
D-Yes	..	...	...	.....	.....	.....	.....	.....	.....	.....	.66	.02
D-No	..	...	...	.....	.....	.....	.....	.....	.....	.....	.....	.16

the standardization of the Bernreuter Personality Inventory it would be expected that each response would be assigned a weight to maximize the correlation of each item response with the total trait score. If this were true, the hypothesized trait would be measured by "Yes" responses, by "No" responses, and by "?" responses. If the "Yes," "No" and "?" responses were measuring the same trait, it should be expected that the correlations of the average trait-response scores would be positive and of such magnitude as to demonstrate the community of the trait being measured. The correlations are positive, but the community is not readily apparent in correlations of the order of +.25 and of +.47. The coefficients of alienation, $k = (1 - r^2)^{\frac{1}{2}}$, are .72, .97, .88 and .75 for B1-N, B2-S, B3-I and B4-D "Yes" and "No" relations, respectively. The alienation coefficients indicate that the assumption that the "Yes" responses and the "No" responses are measuring the same trait (as measured by Bernreuter keys) is hardly tenable. Moreover, the credibility of the assumption of unique or specific trait measurement is further obscured by the fact that some of the correlations of "Yes" and "No" trait-response scores within a trait are exceeded by the inter-trait correlations. For instance, neurotic "Yes" correlates +.68 with introversion "No," whereas neurotic "Yes" correlates +.25 with neurotic "No." Introversion "No" correlates −.70 with dominance "Yes, whereas introversion "Yes" correlates +.47 with introversion "No." It would have been more advantageous to have summed or subtracted "Yes and "No" scores from different traits to have achieved greater *trait* consistency. Thus far, the most favorable aspects of the table have been pointed out. The "?" trait-response scores correlate respectively with the "Yes" and with the "No" trait-response scores +.20 and +.18 for neurotic tendency; +.07 and +.18 for self-sufficiency; +.12 and +.15 for introversion and +.02 and +.16 for dominance. The average of the intercorrelations of "?" with "Yes" and "No" for the four traits is +.14. Clearly, little relationship exists between what "Yes" and "No" measure and what is measured by "?." The traits measured are inconsistent. The "?" responses probably do more to contaminate total trait scores by heterogeneity than they contribute by way of consistency to the trait.

Lorge *op. cit.* has reported that the correlations among the traits are a function of the correlations inherent in the keys. Table II presents additional data which support his conclusion. In Table II are reported the intercorrelations among the weights assigned by the keys

TABLE II.—THE CORRELATION AMONG TRAITS AS EXISTING AMONG THE WEIGHTS ASSIGNED BY "YES" "NO" AND "?" KEYS OF THE BERNREUTER PERSONALITY INVENTORY, AND AS EXISTING AMONG THE AVERAGE RESPONSE-TRAIT SCORE OF ONE HUNDRED THREE EDUCATED ADULTS BY "YES" "NO" AND "?" RESPONSES

Correlations among Bernreuter keys				Correlations among trait response scores			
	S	I	D		S	I	D
N	−.27	.80	−.74	N	−.32	.72	−.78
S	.....	−.08	.43	S	.....	−.30	.53
I	.....	.....	−.53	I	.....	.....	−.65
N	−.25	.80	−.70	N	−.19	.55	−.45
S	.....	−.09	.43	S	.....	−.11	.32
I	.....	.....	−.49	I	.....	.....	−.56
N	−.22	.26	−.18	N	.11	.57	.34
S	.....	−.06	.21	S	.....	−.00	.44
I	.....	.....	−.10	I	.....	.....	.30

separately by "Yes," "No" and "?" keys, and for comparison, the intercorrelations among the average trait-response scores of the one hundred three educated adults. The correlations among the trait-response scores are of the same order of magnitude as those in the keys for "Yes" and for "No" responses. The inconsistency of the trait-response correlations may be related directly to the intercorrelations among the keys.

If the inconsistency has connection with the fact that the keys are not independent, it would be of importance to determine whether keys which are statistically independent result in trait-response scores which are consistent. Flanagan[1] has recently developed two statistically independent trait keys for the Bernreuter Personality Inventory. Through the use of the Hotelling method of principal components he has developed scoring keys which measure two independent components of the Bernreuter Personality Inventory. The intercorrelation among the two Flanagan trait keys are +.04, −.00, and −.04 for the "Yes," for the "No" and for the "?" keys respectively. The traits are statistically independent in the keys. Flanagan has tentatively

[1] Flanagan, J. C.: *Factor analysis in the study of personality.* Stanford University Press, 1935.

named these traits "self-confidence" and "sociability." To avoid error, these two traits shall be referred to as I and II respectively. The same method of analysis employed for the Bernreuter keys will be used to examine the consistency of the new Flanagan keys. The average trait-response score as determined by the "Yes," by the "No" and by the "?" responses of each of the one hundred three individuals was computed by application of Flanagan key I and key II. In addition, the total trait I, and the total trait II, score was computed for each individual.

The intercorrelation of total trait I and total trait II is +.10. Since the keys correlated +.00 (on the average) it indicates that summing "Yes," "No," and "?" weights within a trait spuriously raises the correlation by the mixture of heterogeneous data. It was proposed, however, to examine the consistency of the traits. Table III presents the essential intercorrelations. The average trait response

TABLE III.—THE INTERCORRELATIONS OF THE AVERAGE TRAIT SCORES AS COMPUTED SEPARATELY FOR ALL "YES," FOR ALL "NO" AND FOR ALL "?" RESPONSES OF EACH OF ONE HUNDRED THREE EDUCATED ADULTS BY APPLYING THE FLANAGAN INDEPENDENT TRAIT KEYS DEVELOPED FOR THE BERNREUTER PERSONALITY INVENTORY

where

I-Yes, I-No, and I-? mean the average trait-response score for "Yes," "No" and "?" responses, respectively on application of Key I; and II-Yes, II-No, and II-? have similar meaning on application of Key II.

	I-No	I-?	II-Yes	II-No	II-?
I-Yes	.23	.10	.09	.07	−.07
I-No	...	−.02	.05	−.04	.09
I-?	...	.....	.07	.00	−.07
II-Yes	...	.....	...	.79	.16
II-No	...	.....	...	.....	.14

score as determined by "Yes" responses correlates with the average trait-response score as determine by "No" responses +.23 using key I and +.79 in key II. The correlation of +.23 with its alienation coefficient of .97 can hardly be evidence that the trait measured by "Yes" responses is the same trait that is being measured by "No" responses. Nor could the correlations +.10 and −.02 or of +.16 and +.14 of the "?" trait-response score in each trait with a "Yes" and "No" trait response score be considered as consistent trait measurements. The Flanagan keys are statistically independent, but the scores derived

therefrom lack consistency. The total trait score in any trait for any individual is equal to

$$N_{Yes}(Av._{Yes}) + N_{No}(Av._{No}) + N_{?}(Av._{?})$$

where N_{Yes}, N_{No}, and $N_{?}$ refer to the number of "Yes," "No" and "?" responses respectively, and where $Av._{Yes}$ $Av._{No}$ and $Av._{?}$ refer to the average trait-response scores from these three types of responses. Cognizant of the spuriousness in coefficients of correlation obtained from a part against a total of which it is a part, we, nevertheless, computed the coefficient between the average trait-response and the total trait score to ascertain the relative contribution of each trait-response score to the total. The correlations are:

(I-Yes)	(Total Trait I) = .75	(II-Yes)	(Total Trait II) = .73
(I-No)	(Total Trait I) = .47	(II-No)	(Total Trait II) = .75
(I-?)	(Total Trait I) = .09	(II-?)	(Total Trait II) = .38

Partialling out the influence of the average trait-response scores of I-No and I-? from the correlation of (I-Yes) with Total Trait I, the resulting partial correlation is +.75; partialling out I-Yes and I-? from (I-No) (Total Trait I), +.46; and partialling out I-Yes and I-No from (I-?) (Total Trait I), +.09. The magnitude of the partials for trait I is of the same order as those of the first order coefficients. In trait II, however, such is not the case. The correlation between (II-Yes) and (Total Trait II) partialling out II-No and II-? is +.35; between (II-No) and (Total Trait II) partialling out II-Yes and II-? is +.42; and between (II-?) and (Total Trait II) partialling out II-Yes and II-No is +.41. Clearly, Flanagan's use of the Hotelling component analysis has made a relatively greater contribution by way of "Yes" responses to Trait I than it has to Trait II. The interpretation of the data resulting from the application of the Flanagan keys leads to the inference that the statistical analysis by the Hotelling method has yielded as much a "Yes" trait and a "No + ?" trait as it has yielded traits of "self-confidence" and "sociability." The use of the Hotelling analysis cannot obviate the requirement of logic in the understanding of personality traits. It would have been more profitable perhaps, to have analyzed the Bernreuter traits by the more laborious consideration of the sixty-six correlations of trait-response scores, than the six correlations of total trait scores. Independent keys do not guarantee consistency of trait measurements. Certainly, the modality of response is a variable that must be con-

sidered in personality testing. Statistical rigor may aid in the construction of personality tests, but it cannot be expected to make the traits.

The Bernreuter Personality Inventory has been shown to be lacking in reliability (for individual diagnosis and prognosis), lacking in trait consistency, and certainly lacking in trait independence. Fiat traits are not psychological entities. Since the Bernreuter Personality Inventory lacks the three desiderata of reliability, consistency, and independence of trait measurements, it cannot be validated with any large degree of success.

A COMPARISON OF TWO TYPES OF FOREIGN LANGUAGE VOCABULARY TEST

JOHN M. STALNAKER AND WILLIAM KURATH

The University of Chicago

I. THE PROBLEM

In the past few years, there has been an increasing interest in the use of objective tests for measuring achievement in foreign language work. At the University of Chicago, standardized objective tests have been used systematically in language courses for some years.[1] The advent of the "New Plan," however, with its emphasis on examinations, necessitated the periodic construction of new objective tests for measuring achievement at the various levels of the college program. Some new types or forms of tests have been developed, and old forms have been adapted to the specific objectives of the language work in the college. One of the new types of foreign language tests in vocabulary which have been developed at the University of Chicago is the "Context" test.[2] In this test, the student is required to give the English meaning of certain German words as they appear in context. The experiment described in this article was designed to compare, under controlled experimental conditions, this context test form with an older, well-established form, the best-answer or multiple-choice test, and to determine to what extent the two tests measure the same ability and to what extent these abilities are related to general ability in German.

II. CONSTRUCTION OF THE TESTS

Two German vocabulary tests were constructed, each using the same list of one hundred words; the one form will be referred to as the "Context" test, the other as the "Best-answer" test. One hundred German words were culled from a standard vocabulary;[3] an effort was made to include a normal number of nouns, verbs, adjectives, etc.

---

[1] See, for example, Hagboldt, Peter: "Achievement at the End of the Second Quarter, Measured by the American Council Alpha Test." *German Quarterly*, Vol. I, 1928, pp. 160–169.

[2] Stalnaker, John M. and Eggan, Olive: "The German Requirement for the Ph.D. Degree." *The German Quarterly*, March, 1934, pp. 68–76.

[3] Wadepuhl, Walter and Morgan, B. Q.: *Minimum Standard Vocabulary*. F. S. Crofts and Co., 1934.

and to limit the words to those which are included in the one thousand most frequent words.[4] A short passage of the usual elementary reader type was constructed in which one hundred of the words selected were used. An effort was made not to use the test-words in unusual meanings, and to have the remaining vocabulary of the passage present no unusual difficulties to the student. The one hundred test-words were underlined and each of them repeated at the right of the passage and followed by a blank line on which the student was asked to write the English translation of the underlined word as it appeared in context. He was cautioned to give the correct tense and number of verbs, the correct number of nouns, etc.

The second test was a typical Best-answer test in which the same German words were listed, each followed by five English words; the student was to write the number of the English word which correctly translated the corresponding German word. The words were not arranged in the same order as in the Context test, but care was taken to avoid a marked difference in the inflected forms, except in the case of a few of the commonest strong verbs. The customary precautions were taken to avoid unusual, even though correct, meanings in the case of the correct response, and to have the distractors (incorrect options) within an item equally plausible in meaning and in form. A short excerpt from each of the tests is given in Figs. *A* and *B*.

III. SOURCE OF THE DATA

The two tests were given on February 14, 1935, to one hundred eighty-four students in elementary German courses who had begun their study of the language at the opening of the present school year. All the tests were administered by the authors. The instructors had been asked to announce a test without further comment. Approximately half of the students in each class took the Context test first, followed immediately by the Best-answer; the other half of each class took the Best-answer test first, the Context second. Exactly twenty minutes were allowed for each of the tests. By having alternate students take different forms and by having every other seat vacant, individual work was facilitated.

The instructors had been asked to rate each of their students by a letter grade—*A*, *B*, *C*, *D*, or *F*—indicating their estimates of the

[4] Fourteen of the one hundred words in the test were from the second thousand words.

In the following passage, certain words and phrases are underlined; you will find these words and phrases, followed by a blank line, repeated at the right of the passage. Give the English translation of these words which will fit the context in which the words appear. The English translation should correspond to the German word as to number in the case of nouns, and as to tense in the case of verbs.

EXAMPLE: In der Mostackerstrasze wohnte eine junge Frau, die durch ein Unglück bald nach der Hochzeit ihren Mann verloren hatte.

wohnte *lived*
Unglück *accident*
bald *soon*
verloren *lost*

Ein Mann hatte drei erwachsene Söhne. Diese arbeiteten fast nie, obgleich der Vater ihnen befohlen hatte, ihr eigenes Brot zu verdienen, und böse wurde, wenn sie nicht auf ihn achteten. Der Alte schimpfte oft und versuchte ihnen zu beweisen, dass dieses Leben nicht viel länger dauern konnte, da jeder einzelne von ihnen jetzt mehr verzehrte, als früher alle drei. Obwohl die Lebensmittel damals ziemlich billig waren, schien es ihm unmöglich, dass er sie länger unterstützen könnte.

erwachsene_____
fast_____ obgleich_____
befohlen_____
verdienen_____ böse_____
achteten_____ schimpfte_____
versuchte_____ beweisen_____
dauern_____
einzelne_____ verzehrte_____
ziemlich_____
billig_____ schien_____
unmöglich_____ unterstützen_____

FIG. *A*.

DIRECTIONS: One of the five English words or phrases in each line is a correct translation of the German word. Write the number of the word which you consider to be the correct translation in the blank space.

Example: Mond 1-world 2-month 3-moon 4-mouth 5-word 3

1. bekommen 1-become 2-arrive 3-accept 4-escape 5-receive_____
2. versuchen 1-attempt 2-search 3-request 4-conceal 5-visit_____
3. beweisen 1-prove 2-inform 3-bleach 4-direct 5-advise_____
4. verdienen 1-administer 2-serve 3-hire 4-dilute 5-earn_____
5. Pflicht 1-situation 2-pledge 3-escape 4-duty 5-pride_____
6. Angst 1-narrowness 2-fear 3-stallion 4-proximity 5-autumn_____
7. ziemlich 1-timely 2-docile 3-cozy 4-rather 5-angry_____
8. erleben 1-survive 2-vivify 3-elevate 4-experience 5-praise_____
9. schliesslich 1-enclosed 2-finally 3-secure 4-rapidly 5-attainable_____
10. Vetter 1-cousin 2-weather 3-parent 4-gambler 5-sprinkler_____
11. gehören 1-belong 2-hear 3-raise 4-acquire 5-explode_____
12. schlecht 1-striking 2-slight 3-badly 4-rainy 5-belligerent_____
13. Kleidung 1-clotting 2-ascent 3-renovation 4-apparel 5-accompaniment_____
14. allerlei 1-various 2-primary 3-everything 4-in addition 5-oldish_____
15. etwa 1-something 2-dual 3-perhaps 4-closely 5-broken_____

FIG. *B*.

student's general attainment in German at the time the tests were administered.[5] All ratings were obtained before the tests were given. Scores on the intelligence tests given at the time of entrance to the University were also available for one hundred twenty-three of the students.

On the day following the examination, each student was asked to indicate in writing which of the two test forms he considered the fairer test of his ability in the language and to give his reasons.

Mr. Kurath personally scored all of the Context tests. They were scored very strictly; unless a word was translated so as to fit into the context in which it appeared, and unless it was of the correct number and tense, it was counted as incorrect. The Best-answer test was scored by key, the total score being the number of correct responses. No correction for guessing was made.

IV. RESULTS

The gross results of the tests are given in Table I. The absolute scores on the Best-answer test are significantly higher than those on the Context test. This difference is usual between recall and

TABLE I.—THE NORMS OBTAINED FROM THE ONE HUNDRED EIGHTY-FOUR STUDENTS WHO TOOK THE TESTS

	Best-answer	Context	Difference
Mean	47.01	32.73	14.28
Sigma	17.79	18.26	−0.47
First quartile	32.70	17.70	15.00
Median	46.30	31.70	14.60
Third quartile	60.50	45.50	15.00

recognition tests. The dispersion of scores, however, is approximately the same on the two tests. In other words, by adding a constant—approximately fifteen—to the mean and quartiles of the Context scores, the measures of the Best-answer test will be obtained. The insignificant differences between the means and medians suggest that the distributions are not skewed.

The reliability of the tests was computed by the usual split-half method. The reliability of the twenty-minute Best-answer (recogni-

[5] For this cooperation, the authors are indebted to Miss Viola Manderfeld and Messrs. Charles Goetsch, H. A. Basilius, and K. M. Grubb.

tion) test is .95; that of the twenty-minute Context (recall) test is .97. The difference of the two reliabilities is significant.[6] These reliabilities are unusually high for a short test which has been subjected to no preliminary statistical analysis. All the one hundred eighty-four students who took the test were from the same course in German. The reliabilities were felt to be highly satisfactory and to justify the strictness of the scoring procedure adopted.

As has been stated, approximately half of the group, ninety-eight students, took the Best-answer test first, and the remaining students, eighty-six in number, took the Context test first. Whether a student took the tests in one order or the other was determined by the way the students happened to seat themselves in the room. The hope was that the two groups would be of about the same ability. Table II

TABLE II.—A COMPARISON OF THE GROUP WHICH TOOK THE BEST-ANSWER TEST FIRST WITH THAT WHICH TOOK THE CONTEXT TEST FIRST

Instructors' estimates	Group which took	
	Context first (86)	Best-answer first (98)
Mean	5.24	4.67
Sigma	1.52	2.09
Difference of means / Standard error of difference	$\frac{.57}{.277} = 2.06$	
Intelligence Scores		
Mean	63.17	52.74
Sigma	28.50	29.82
Difference of means / Standard error of difference	$\frac{10.43}{5.16} = 2.02$	

shows the data for the two groups. The group which took the Context test first was the superior group on the average, as judged by the ratings of the instructors and by the intelligence ratings.

Because certain logical reasons might be advanced why the taking of the Context test would aid in taking the Best-answer test, a compari-

[6] The difference of the z coefficients is $0.31 \pm .105$. The ratio is thus 2.95. See Fisher, R. A.: *Statistical Methods for Research Workers* (third edition). London, p. 170.

son of the scores of the two groups was made. Inasmuch as the group which took the Context test first was the superior group, they would be expected for this reason alone to score higher. Actually, the mean total score (sum of the two tests) of those who took the Context test first was 86.74, and of those who took the Best-answer test first, 74.39. This difference of 12.35 has a standard error of 5.02. The ratio of the difference to its standard error is 2.46, which suggests that the difference is a significant one. Table III gives the differences on the two separate tests.

TABLE III.—THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE SCORES OF THE GROUP WHICH TOOK THE TESTS IN ONE ORDER AND THAT WHICH TOOK THEM IN THE OTHER ORDER

	Mean scores of the group which took first the		Difference of means	Standard error of difference	Ratio of difference to SE
	Best-answer test	Context test			
Best-answer test...	40.43	54.51	14.08	2.43	5.8
Context test.....	33.20	32.19	1.01	2.67	0.4

In order to rule out the influence of the order in which the tests were taken, standard scores were computed separately for the two groups. The standard scores for the group which took the Context test first were based on the means and standard deviations obtained from that group; the standard scores for the group which took the Best-answer test first were likewise computed from the constants obtained from that group. All subsequent data reported in this paper are based on these standard scores.

The chief question of interest is the relative validity of the two forms. Our criterion of validity—instructors' estimates—was a fallible one, but the best available. No measure of the reliability of these estimates was made. The correlation of the Context test and the instructors' estimates was .629; that of the Best-answer test with the estimates was .595. The difference between these two correlations is not significant.[7] The two tests, then, are, as nearly as we can tell from these data, equally valid. In each case, the validity coefficient of about .60 is, it may be noted, relatively high for a twenty-minute

[7] The difference of the z coefficients is .055 ± .111. *Cf.* footnote 6.

test devoted to only one aspect of German—vocabulary. The estimates were based on general ability in the language.

If the two tests are about equally valid, and equally reliable, the question as to their relationship to each other naturally arises. The correlation of the Context and the Best-answer tests was .911. The reliabilities of these tests were so high that any correction for attenuation due to unreliability will be slight.[8] The corrected coefficient is, however, .951. In other words, the two tests seem to measure very nearly the same thing.

The Best-answer test correlates with intelligence .306. The Context test correlates with intelligence .342. The difference here is not a significant one.[9] The two tests seem to be about equally related to intelligence. The correlation between the estimates made by the instructors and the intelligence ratings was .487.

On the day after the test was given, the regular instructors asked the students to indicate on a slip of paper which of the tests they thought was the fairer test of their knowledge of German and to give reasons for their answers. One hundred forty-nine papers were received. These were classified as favoring the one form or the other or as being neutral. Fifty per cent favored the Context test; forty per cent favored the Best-answer test, and ten per cent were neutral.

The reasons given for the preferences were in general important ones. The Best-answer (recognition) test was preferred because one got a higher absolute score on it, one could answer a greater amount of it within the time limit, each question was a unit and independent of other questions, guessing was limited to one of five (whereas in context it is unlimited!), it is subject to more accurate grading. The Context (recall) test was preferred because it tested the more vital thing, the ability to read and not mere memory of words, it reduced possibility of guessing, it was the natural test, not an artificial pedagogical device, and it was the better test of what the students are attempting to learn to do.

V. SUMMARY AND CONCLUSIONS

Two highly reliable twenty-minute tests in German vocabulary were constructed—a best-answer recognition test and a context

[8] The correction should probably be larger than it is. See Remmers, H. H.: "A Possible Experimental Error in Determining the Overlap of Two Correlated Variables." *The Journal of General Psychology*, Vol. IX, 1933, pp. 459–461.

[9] The difference of the z coefficients is .040 $\pm$ .129. *Cf.* footnote 6.

recall test—and administered to one hundred eighty-four students in "German 102," elementary German. The results show that the Context test is slightly more reliable and is preferred by a slightly greater number of the students as the fairer test. The two tests correlate with intelligence and with instructors' estimates to about the same extent. The Best-answer form is easier to score. The results show that the two tests measure very nearly the same ability. Which test should be used in a given situation would depend on factors other than those which were studied in this experiment.

The data gave further evidence that in the actual test situation, recall tests and recognition tests measure essentially the same ability, and that the lucky guessers on the recognition test are also the fortunate recallers on the recall test.

EMOTIONAL FACTORS IN READING DISABILITIES IN SCHOOL CHILDREN[1]

SIMON H. TULCHIN

New York City

Although the past two decades have witnessed considerable progress in the recognition of individual differences in school children, very little has been accomplished in the practical problem of treating these differences. There is always an appreciable lag between the discovery of new principles and methods in any field and their practical utilization. This lag is especially pronounced in the field of education. It is also essential to remember that even when new and more desirable methods of instruction are introduced they are concerned primarily with the majority of pupils and fail to make adequate provision for those who show marked deviation from the average. Frequently even progressive schools organized to give individual instruction to their pupils commit the error of utilizing the same methods of instruction with all children, failing to make adequate provision for individual differences.

More recently the growth of the Mental Hygiene movement and the development of Child Guidance clinics which coordinated the techniques of psychiatrist, psychologist, and social worker have brought to the foreground the emotional problems of the child. We are beginning to consider not only problems of teaching but also problems of learning. How does the child feel when trying to learn? Is the particular method of instruction used adapted to his special needs? Is the teacher aware of the emotional factors which may be facilitating or inhibiting the learning process? One can readily see that emotional factors are especially important where children presenting specific disabilities are concerned.

It is of interest to note that the first studies of reading disabilities were made by physicians whose primary concern was with the physical aspects of their cases. Psychological and emotional factors came to be considered much later.

It is not the purpose of this paper to evaluate the several theories proposed to account for the causes of reading disability. The reading

[1] Read before the Joint Meeting of the Mental Hygiene Department of the New Jersey State Teachers Association and the New Jersey Association of Psychologists, at Atlantic City, November 10, 1934.

process involves the coordination of many component factors—auditory, visual, motor, etc.—and an inability to learn to read may be the result of a failure in any one of these fields. We are not concerned with the cases in which mental deficiency in itself accounts sufficiently for the inability. The failure to learn to read may also depend primarily upon emotional factors. These are sometimes extremely difficult to analyze, as so many of the cases seen by the clinical psychologist or psychiatrist are frequently brought to his attention after a considerable lapse of time since the disability was first manifest. The child's experience during the first few reading lessons may be so charged emotionally as to color all his subsequent reactions and determine his resistance to reading. In such a situation no other reading disability need be present. His emotional reaction to reading may even be determined by conditions in the child's environmental set-up long before any reading lessons are attempted. On the other hand, the child with a specific reading disability which is non-emotional in character may develop various personality and behavior difficulties as a result of his school maladjustment. This possibility is clearly recognized by several workers in the field.

General instability, flighty attention, resistance to authority, feelings of inadequacy, infantile behavior, sibling rivalry, oversensitiveness to criticism and other emotional factors singly or in combination may interfere with learning to read. These and other emotional reactions may in turn trace their origin to a reading disability based upon other than emotional factors. Thus Orton writes:[1] "Our studies of children with reading disabilities have also brought to light certain other aspects of the problem which are of educational importance . . . Among these were notably the effect of this unrecognized disability upon the personality and behavior of the child. Many children were referred to our clinic by their teachers in the belief that they were feeble-minded, others exhibited conduct disorders and undesirable personality reactions which upon analysis appeared to be entirely secondary to the reading defect and which improved markedly when special training was instituted to overcome the reading disability." And again:[2] " . . . From these emotional stresses we have seen the development of several reaction patterns. Some children

---

[1] Orton, Samuel T.: "The 'Sight Reading' Method of Teaching Reading." (Complete bibliographical references will be found at the end of the paper.)

[2] Orton, Samuel T.: "An Impediment in Learning to Read."

become apathetic and indifferent to school. Others build a marked feeling of inferiority, while a few develop an antagonism toward their teachers and not infrequently become disciplinary problems. The emotional load seems to be particularly troublesome when a younger brother or sister is making more rapid progress." Orton goes on, citing a case of a boy with a reading disability who "showed very little evidence of strain until his six year old brother entered school and began to make headway in reading." A vivid picture of maladjustment is presented by Newell[1] who writes: " . . . The mysterious failure to learn to read on the part of an apparently intelligent child causes anxiety and misunderstanding. The parents can account for the difficulty only on the basis of obstinacy or mental defect and they resort to reproaches, nagging and even punishment. The child becomes bewildered and resentful and is antagonistic toward all attempts to help him. The teacher who has thirty or forty other pupils can only put him in her slowest group. If she does not understand his difficulty she cannot help him. The unhappy plight of the frustrated non-reader sometimes contributes to undesirable forms of behavior as a means of obtaining the prominence which he cannot gain by acceptable means."

The interplay between the reading disability and the emotional factors, especially in families of unstable stock with high academic standards, is further described by Hincks:[2] " . . . The emotional difficulty is increased by the emotional traits of the child, which in turn are augmented by the disapproval and worry of excitable parents. A bad state of tension develops, in which the reading adds to the general nervousness of the child's condition, and this condition adds further inhibition to the learning process, and irritation to the parent, so that a general family and school maladjustment occurs."

It is not always easy to determine whether emotional and personality factors interfering with learning to read are the primary causes for the disability, or are themselves merely the result of the disability. Of course, both emotional and specific reading disability factors may be present in the same child, with each set of causes having an independent origin. One valuable clue is offered in the response of the child to remedial efforts. It seems that the more primary the emotional factors the greater the stumbling block in

---

[1] Newell, Nancy: "For Non-Readers in Distress."

[2] Hincks, Elizabeth: *Disability in Reading in Relation to Personality.*

treatment. Also when emotional factors seem primary, disability and general lack of progress in other subjects as well as in reading is more likely to occur.

In carrying remedial treatment work with children who have difficulty in learning to read, one is continually impressed with their over-sensitiveness to criticism, their lack of self-confidence and feelings of insecurity, their short attention span, their general state of tension, and their resistance to remedial effort. In discussing the general treatment procedure Monroe[1] stresses especially the need for putting the child at ease, of helping him to get the relief which comes from going back to simple tasks, of praising liberally for success and overlooking errors, for individual instruction, and for modification of methods used to suit the needs of each child. In a more recent monograph Monroe[2] stresses the need for adequate motivation which comes largely through the recognition of the child's own success.

This emphasis on interest and desire to learn is also made by Lord[3] who writes: " . . . The receptive attitude on the part of the child must be supported by constant encouragement to the point where the feeling of inadequacy with its inhibiting effect is replaced by the belief in his own ability. The pedagogic methods are secondary to this . . . " Hollingworth[4] argues that no single method of remedial instruction established exclusive claim and adds: "In fact no investigator has established his or her method as the only method of successful approach to particular cases by excluding other methods through experimental teaching." She adds that the best methods consist in the utilization of all avenues of approach.

Experience has shown, then, that when children fail to learn to read when exposed to the usual classroom procedure an attempt should be made to analyze the reasons for their failure as soon as the difficulty becomes evident. The optimistic faith of some teachers that the disability will be "outgrown" is fraught with real danger. The study of the child should include not only diagnostic reading tests which are exceedingly helpful in giving us a picture of the child's present reading status and frequently invaluable assistance in determining the causes for the disability, but also intelligence tests, both

---

[1] Monroe, Marion: *Methods for Diagnosing and Treatment of Cases of Reading Disability.*

[2] Monroe, Marion: *Children Who Cannot Read.*

[3] Lord, Elizabeth: *Special Disabilities in Learning to Read.*

[4] Hollingworth, Leta: *Special Talents and Defects.*

verbal and performance, achievement tests of all school subjects, a physical examination with special eye and ear tests when indicated, eyedness and handedness tests, a check up of the educational methods employed by the school in teaching reading, and an analysis of the child's personality and emotional make-up and environmental set-up with special emphasis on their bearing on the reading situation.

The remedial program must be worked out in relation to the findings disclosed by the study and will depend upon the special factors in each individual case. The very nature of the problem precludes any rigid adherence to a single program or method, and a great deal of flexibility is demanded. The child himself frequently gives clues which can be utilized to advantage. Changes may be indicated at each tutoring session, especially during the initial stages of therapy when much depends upon overcoming the resistance of the pupil, his lack of confidence and frequently his discouragement which resulted from previous failure. The intensity of the treatment and the period of time required obviously depend upon the nature and severity of the problem. It frequently happens that remedial work is terminated by the parent or by the school as soon as first indications of progress are made. Treatment should not terminate before the child's achievement in reading as measured by tests is in keeping with his progress in other subjects and with his general intellectual endowment.

The following excerpts from case records briefly illustrate some of the problems presented by children who have difficulty in learning to read.[1]

CASE I

Peter was referred as a reading disability case by the school. He is a slow oral reader whose difficulty dates back to the beginning of his reading experience. He shows no interest in reading at home and will never pick up a book to read for pleasure. If a Popular Science magazine is about he will go through it with great interest, and can tell all about the inventions, but will not read aloud. In school he can read silently but not orally. He seems sensitive, blushes and cannot read. He is right handed. Peter was in 5*A* at

[1] The author wishes to thank Dr. Lawson G. Lowrey, formerly Director of the Institute for Child Guidance, New York City, for permission to use Institute case material and Dr. David M. Levy, New York City, for permission to use his private case records.

the time of referral. He skipped grades 2*A* and 4*B*. When in the second grade the teacher told mother that he did not read and seemed very sensitive about it, often crying. If he couldn't be the best it hurt him. His present teacher reports that Peter is good in all subjects except oral reading. He blushes and seems pained when called on to read. Both parents tease him about reading. His father, who is a fairly successful man, says, "I could never read. How do you expect Peter to read?"

Peter, who was nine years and two months old when tested, made a mental age of thirteen years nine months on the Stanford Binet examination. This gave him an IQ of one hundred fifty, placing him in the exceptionally superior group. His average age on a group of performance tests was nearly two years above his life age. On the Stanford Achievement test he scored one half year below his grade placement in reading, while in arithmetic he scored one year above his present grade. Diagnostic reading tests showed him to have a reading disability. He made numerous reversals of letters and words as well as other errors.

No other problems except the over-sensitiveness which related directly to the reading disability were noted. The reading disability was explained to the parents as well as to the boy and special tutoring recommended. While the parents readily agreed, it took some persuasion to get the cooperation of Peter. It seems that this exceptionally superior boy found it extremely difficult to admit, even to himself, any problem in relation to his school work. It was of course his general intelligence and determination that made it possible for him to carry his school work and even skip some grades. With his high intellectual level he was able to get the content of his lessons when he did not have to compete in oral reading.

A book which dealt with popular science was selected for him. Once his cooperation was secured it took only two months to achieve very satisfactory results. A note in the record some six weeks after tutoring began reads: "Peter came in for tutoring. He is enthusiastic about reading he is doing at home and is enjoying it. He is quite proud of his grades on his last report card, which his teacher describes as a great improvement." Five weeks later both mother and the school report remarkable improvement in reading. Peter now reads voluntarily at home. At the end of the school term Peter was allowed to skip a grade. "He now reads with enthusiasm" says a note in the record and adds that the school has noticed a splendid improvement

not only in reading but in general self-confidence which Peter has acquired. A re-examination done a year later showed no reversals and no other errors. Mother reports that Peter was so absorbed in a book that it was difficult to get him to stop, and that he read many books since he discontinued contacts at the Institute. His school work is excellent. Another follow-up was made two years later. The same enthusiastic reports are found in the record. He was soon to enter High School and was interested in engineering.

In this case the school recognized the reading problem as early as the second grade but did not undertake any special treatment, apparently because of the child's exceptional intelligence. The parents merely teased him instead of providing the needed help. Undoubtedly if allowed to continue much longer without the special help more serious emotional problems would have resulted.

In contrast to this case we now turn to Case II.

CASE II

George was referred because of persistent truancy from Continuation School, refusal to work, extreme sensitiveness and instability and because he was unmanageable at home. At the time of examination he was sixteen years eight months of age. He was discharged from the beginning of the sixth grade soon after he reached his sixteenth birthday and at first sent to a Probationary School. During his school career he changed schools four times. His marks for conduct were *A* in grades 1*A* to 3*B*, *B* for grades 3*B* to 5*B*, and *C* in 5*B*. His school record shows that he remained in

Grade	Terms	Grade	Terms
1*A*	2	3*B*	1
1*B*	3	4*A*	1
2*A*	2	4*B*	1
2*B*	3	5*A*	1
3*A*	3	5*B*	3

In ten years he made five grades. A note in the history states that George liked the Probationary School because he had wood and electrical work there. The teacher there did not find him troublesome at all. George disliked all other schools.

The psychological tests gave him a mental age of thirteen years five months and an IQ of eighty-four on the Stanford Binet examina-

tion. He failed the ten year vocabulary but showed a good memory for digits and passed the clock test in year fourteen and the digits code and Boxes tests in year sixteen. On the Army Alpha test he scored thirty-three points, which is equivalent to a mental age of about twelve years. On a series of four performance tests he made the highest possible scores on all tests, which gave him a mental age above eighteen years. Thus he made perfect scores on the Porteus, Healy P. C. II, Ferguson Form Boards, and the Kohs Block Design test. He also made a perfect score on the Stenquist Mechanical Assembly test—a score equalled by only one per cent of adults. During the performance tests he appeared more spontaneous and confident in himself. He talked of his interest in radio telegraphy and electrical devices. The psychiatrist's notes describe George as presenting a marked picture of inferiority and inadequacy. At no time did he achieve success.

During the Alpha examination George remarked that he could not read very well. Because of this and because of his school history and the marked discrepancy between the results of the verbal and performance tests diagnostic reading tests were given. The results showed typical reversal, errors in consonants and vowels, etc. which are found in reading disability cases. Here was a boy who was kept in school for ten years and finally discharged the moment his age allowed it without the slightest recognition of his reading disability or special ability with concrete and mechanical material. While at first compliant, as is shown by his conduct marks of *A* in the first three grades, which took seven years to complete, George finally became discouraged and all the other problems followed in relatively short order. His regular progress in 3*B* through 5*A* is not accounted for as no information was available. It would be of interest to know what happened, or is it likely that the teachers merely got tired of him and pushed him on! Is it any wonder that George became defiant and unmanageable? Unfortunately the economic situation of the family was such that it was imperative that George should go to work. Unfortunately also George's resistance toward anything that had to do with school was so great that remedial work in reading was practically impossible. The economic situation in the country at the time made it difficult to secure the kind of job for George which would make use of his special abilities. A job was secured for him temporarily. He showed good stability and interest in his work and when last heard from two years later he still held his job.

It is interesting to speculate what George's career might have been had the school discovered and treated his reading disability when it was first manifest.

CASE III

In the third case the reading disability is closely tied up with a sibling rivalry situation. The boy who is nine years four months of age makes a superior rating on the Stanford Binet examination. His mental age is ten years nine months, IQ one hundred fifteen. The IQ is considered somewhat minimal because of the reading disability. His average age on a group of performance tests is the same as the Binet MA. On the Stenquist Mechanical Assembly test he shows exceptionally superior ability. Diagnostic reading tests disclose a reading disability which has been discovered as soon as he entered school and special help given at irregular intervals with little success. John was in the fourth grade at the time of study. Achievement tests showed him to be one year below his grade in arithmetic and two years below in reading. He is right handed and right eyed. When asked, during the test, whether he uses his left hand for throwing a ball he replies: "I use my left hand for only one thing—to punch my brother."

The sibling rivalry situation and its relation to the reading problem is brought out during interviews with the psychiatrist. John is constantly derogatory to the brother, dwelling on all his weaknesses and when challenged to mention one thing about his brother that is good he can only think of: "He can read and write better than I can." Challenged to say one other thing he replies: "That is too hard—I'm stuck in the mud." The brother, who is nearly two years older, was able to read at a time when John did not yet enter school. When asked to remember as early as he can a scene in which the brother is reading to the mother he tells of seeing the mother and brother sitting together, himself a distance away, and describes his desire to "do something" to get her attention. In a phantasy he fights with the brother and actually gets him away. He goes into detail as to how he would lick the brother, what blows he would administer, etc. His interest in mechanical things seems directly associated with a desire to excel in a field of his own.

A short period of tutoring simultaneous with psychiatric treatment of the sibling rivalry situation cleared up the reading disability and greatly improved the rivalry problem. A follow-up report from the

mother one and one-half years later informs us that John has made considerable improvement in reading and now picks up books to read by himself. He was promoted regularly and does good work in school in all subjects. It seems quite reasonable to believe that the failure of previous tutoring attempts was largely due to the lack of recognition of the emotional problems created by the sibling rivalry situation.

CASE IV

This is a case of a ten year old boy with an IQ of ninety-four on the Stanford Binet examination. The rating was considered somewhat minimal. His average age on a group of performance tests was the same as the Binet MA. No history of left handedness. He was unable to score on any of the Stanford Achievement tests which involved reading but scored up to grade on the arithmetic computation test. The diagnostic reading tests, which could only be given in part, disclosed many reversals, much confusion, etc. He could not write his last name. His entire behavior during the tests was extremely infantile. It was difficult for him to give sustained effort on any verbal material. He seemed sensitive to failure.

The family history shows much marital difficulty which went on for several years and terminated in a divorce. James, who was the mother's favorite, was exposed to these difficulties nearly all his life. He appeared always in a state of anxiety.

Progress in both the tutoring sessions and psychiatric treatment was slow. Remedial work was definitely hampered by the boy's infantile behavior, lack of motivation, narrowed contact with reality, and deeply rooted emotional problems. One may definitely say in this case that without psychiatric treatment progress in reading cannot be made. While after five months of treatment some progress has been shown, much remains to be accomplished. Neither have the emotional problems been solved. Since this case is still under treatment no final report can be given at this time.

The school can do much, not only in selecting children with special disabilities for study and treatment but also in carrying on remedial work providing the problems presented by the cases selected are not in need of specialized techniques and experience.

The following excerpt describes a social worker's visit to the school in behalf of a reading disability case who was brought in for study because the little girl from three weeks old had been sensitive and cried easily. "The principal recalled the family, remarked on

what 'charming tow-heads' the children were, that they caused no trouble, hardly allowing the worker a chance to explain the object of her visit. She said that all children had difficulties, the teacher knew them all, she promoted everyone if possible, etc. Before worker realized it, patient's present teacher was summoned and when she could not remember Mary, she was waved out with 'She will, my dear, she will.' Worker finally was able to explain to the teacher how sensitive Mary was and asked for cooperation, assuming that she would take special interest in the child. Worker suggested that at present her reading difficulty be minimized and her accomplishments in other subjects be remarked upon. When worker returned patient was in the principal's office. The principal was examining her teeth, saying 'Why, what an idea—you can read. Why, how many years bad are you, etc.,' cajoling her. Patient did not know how to take her and was bravely fighting tears. When she saw worker she ran toward her, clinging to her. Worker tried to smooth over the situation, talking about Mary's vacation, how attractive she looked, etc." Worker then terminated the interview with the principal thanking her for her interest. In the concluding sentence worker tells that she felt "the contact at the school was an unfortunate one because she had hoped that Mary would not learn of worker's visit."

The above incident is mainly of interest because it demonstrates how, even with best intentions, a teacher or principal who lacks sufficient information about the emotional problems of a child may err in her approach to the child's difficulties.

The teacher's overcrowded schedule and numerous duties do not always make it possible for her to give the specific help which so many of her pupils may need but she is in an advantageous position to recognize the individual who presents specific problems and may contribute much to a child's adjustment in later life by early recognition of these difficulties. Adequate treatment involves a clear recognition of all the factors that determine the problem as well as the necessary therapeutic technique.

BIBLIOGRAPHY

Blanchard, Phyllis: "Reading disabilities in relation to maladjustment." *Mental Hygiene*, Vol. XII, October, 1928, pp. 772–788.

Bronner, Augusta F.: *The psychology of special abilities and disabilities.* Boston: Little, Brown and Co., pp. 92–110.

Dearborn, W. F.: "Nature of special abilities and disabilities." *School and Society*, Vol. XXXI, May 10, 1930, pp. 632–636.

Gates, A. I.: *The improvement of reading.* Macmillan, 1929, pp. 22, 23.

Gray, W. S.: *Remedial cases in reading: their diagnosis and treatment.* Supplementary Educational Monographs. Department of Education, New York City, 1922, No. 22.

Hincks, Elizabeth: *Disability in Reading in Relation to Personality.* Harvard Monographs in Education. Cambridge University Press, Whole No. 7, Vol. II, 1926.

Hollingworth, Leta S.: *Special Talents and Defects.* Macmillan, 1931.

Lord, Elizabeth E., Dearborn, W. F., and Carmichael, L.: *Special Disabilities in Learning to Read.* Harvard University Press, 1925, $1.00.

Meyers, Garry C.: "Schoolroom hazards to the mental health of children." *Mental Hygiene,* Vol. XII, January, 1928, pp. 18–24.

Monroe, Marion: *Methods for Diagnosing and Treatment of Cases of Reading Disability.* Genetic Psychology Monograph, Vol. IV, October, November, 1928.

Monroe, Marion: *Children Who Cannot Read.* University of Chicago Press, 1932.

Newell, Nancy: "For Non-Readers in Distress." *Elementary School Journal,* Vol. XXXII, November, 1931, pp. 183–195.

Orton, Samuel T.: "An Impediment in Learning to Read"—a neurological explanation of the reading disability. *School and Society,* September, 1928, pp. 286–290.

Orton, Samuel T.: "The 'Sight Reading' Method of Teaching Reading," as a source of reading disability. *Journal of Educational Psychology,* Vol. XX, February, 1929, p. 135.

THE OUT-OF-SCHOOL ACTIVITIES OF WELL-ADJUSTED AND POORLY ADJUSTED ELEMENTARY SCHOOL PUPILS

MARTHA CRUMPTON HARDY

Elizabeth McCormick Memorial Fund

For a number of years the writer has been carrying on a series of observational studies on a group of pupils in their progress from the third grade into the high school. This article summarizes some of the findings with respect to the relationship between the things which engaged the out-of-school hours of these pupils, during the elementary school years, and certain patterns of behavior exhibited in the school and neighborhood situations.

The pupils were selected mainly on the basis of the willingness of the parents to cooperate in a long-time program[1] and some assurance that they would remain in the locality. With respect to intelligence and home conditions, the pupils followed can be said to represent a sample of the public school population.

Information relative to the out-of-school activities was given yearly by the child and his teachers, and in some instances by his parents. The procedures for collecting the data included tests, checking devices, rating scales, questionnaires, and narrative observations. The activities herein reported concern the things the child called play in which he liked or disliked to participate; some organized forms of recreational and educational activities; attendance at the movie; the reading of books; and remunerative employment.

The sources of the behavior data were teacher ratings and observations, pupil interviews, and parent reports. A schedule consisting of a short questionnaire and a rating scale for twenty-seven traits was given to the teachers four weeks after the beginning of each semester. The ratings during the elementary school years were treated as a whole, each teacher's being given equal weighting with the others. Patterns were determined by the extreme deviates as considered from several points of view:

1. As to, general classroom behaviors which were said to be most unlike those exhibited by the majority of the pupils.

[1] These data form a part of a general investigation on Health Education and Child Development, directed by the late Carolyn Hoefer, which the Elizabeth McCormick Memorial Fund has been conducting in the public schools of Joliet, Illinois.

2. As to, special traits and attitudes in which wide differences were noted, *e.g.*, in attitudes toward classmates, toward the school authorities, and toward assigned tasks.

3. As to, general behaviors reported as varying markedly from those observed among children of the same sex. The boys were judged in terms of the scores for boys, and girls in terms of the behaviors reported for girls.

Then, from parent and pupil interview records, evidences of desirable traits were shown by a high degree of social recognition from the group, and of undesirable traits by excessive fears, frequent outbursts of anger, and withdrawal tendencies.

The extreme cases included seventy-three boys and seventy-one girls in whose behavior, during a four-year observation period, definite indications of incipient maladjustments were noted, and thirty-three boys and forty-six girls who appeared to be building up very desirable patterns of behavior. There were five per cent of the boy population and seven per cent of the girl population who were classified as poorly adjusted from the standpoint of both their general behavior in the school environment and of special types of personal problems.

Instead of differentiating these cases into two classes as representatives of satisfactory and unsatisfactory behavior, the findings are presented separately for each of the several methods of judging personal adjustments. Interpretations have been based upon the uniformity in the trend of these results, as well as on the size of the critical ratios of the mean differences.

On the whole, the poorly adjusted pupils were from a somewhat lower socio-economic level than the average pupil in the investigation and from a markedly lower level than the well-adjusted pupils. Since this factor of home condition would seriously complicate the issues of our problem, it was necessary to control it. The procedure followed was that of making the distributions of the poorly adjusted proportionate to the well-adjusted with respect to occupational status of the fathers, cultural level of the parents (as judged by amount of schooling) and the physical aspects of the home and neighborhood.

The most outstanding difference between the behavior deviates related to their general intelligence. Even when the home factor was controlled, the well-adjusted pupils scored as far above the average of the observed population on the intelligence test as the poorly adjusted scored below the average. The scales used to measure socio-economic level were the Whittier Scale for Grading Home Conditions and Barr's scale for occupational intelligence. The intelligence quotient was based on the average from yearly tests on the Stanford Revision.

The sex distribution of the behavior deviates was as follows:

	Boys *N*	Girls *N*	Boys and girls *N*
General adjustment.			
Well-adjusted	17	16	33
Maladjusted	17	14	31
Special traits.			
Well-adjusted	4	21	25
Maladjusted	24	7	31
Best liked	15	23	38
Fearful	9	34	43
Quick-tempered	28	14	42
Withdrawing	19	17	36
Total population	215	194	409

RESULTS

Since the well-adjusted pupils had higher intelligence scores than the other pupils, some variations in certain types of their interests would be expected. However, there was no indication that the well-adjusted representatives were either more or less versatile in their play interests; more or less proficient on the playground; nor, that they participated either more or less extensively in forms of organized activities than their less satisfactorily adjusted classmates. There, likewise, was no suggestion of a consistent difference in the number who were remuneratively employed after school hours and over week-ends.

With respect to reading interests, while the well-adjusted reported a larger amount of leisure-time reading and expressed greater enjoyment in this type of recreation, when the cultural level of the home was similar, the differences, regardless of a wide divergence in pupil intelligence, were too small to eliminate chance variations. At the yearly interviews with the psychologist, comments on the child's expressed preference for reading content were recorded. The descriptions given during the elementary school years were combined and classified independently of the questionnaire material on actual reading done. Here again, an examination of the data showed no consistent differences which appeared to be associated with the character of the child's behavior.

The only activities in which significant variations between the extreme cases were indicated related to their movie attendance, the size of their play units, and to some aspects of their play behavior. An interesting possibility was suggested as to differences in the type of organizations to which these pupils gave their "free hours." The comparative findings presented have been restricted to those suggestive of differences between the behavior deviates and the average child in the study. Tables I and II show the extent to which the poorly adjusted pupils appear to differ from their better adjusted classmates.

Attendance at Movies.—The movie attendance record is limited to data collected when the pupils were in the fifth grade. With the exception of the representatives of withdrawal traits, the pupils evincing consistent signs of poor adjustment reported more frequent attendance at the movies, and those exhibiting desirable patterns of behavior less frequent attendance, than the average pupil in the study. It can be stated with a reasonable degree of confidence that pupils who were meeting with difficulties in making adjustments in the school environment attended the movies more regularly and more frequently than pupils, from a similar socio-economic level, whose classroom and playground adjustments were called satisfactory. The critical ratios in standard error units were in this instance greater than 3.0.

Among the quick-tempered-pugnacious group, sixty-eight per cent went regularly at least once each week and thirty-two per cent went at least twice weekly. Their record was significantly different from that of the average child from the same level of intelligence and from similar home conditions. (Critical ratios 3.4 and higher.) But there were twice as many boys as girls in this group and, since all studies of movie interest have indicated a higher frequency of boys at movies than girls, it was necessary to check further these results. A contrast with the boy population showed that frequent attendance (two or more shows weekly) was twice as common among the quick-tempered as it was among boys in general, which would tend to establish the trend of the results.

The excessive fear cases were mainly a girl group. Among them, attendance as often as twice a week was four times as common as it was among the satisfactory behavior cases, and three times as prevalent as among the girl population.

On the other hand, the shy, sensitive children reported less regular attendance at the picture show than the average child. Reference

TABLE I.—SHOWING DIFFERENCES BETWEEN WELL-ADJUSTED AND POORLY ADJUSTED PUPILS WITH RESPECT TO INTERESTS IN MOVIES AND IN ORGANIZED FORMS OF ACTIVITIES

Activities	Behavior groups									Observed population
	General adjustment			Special traits		Best liked	Fearful	Quick-tempered	Withdrawing	
	Well-adjusted[1]	Maladjusted[1]	Critical ratio $\left(\frac{\text{Diff.}}{\text{SD diff.}}\right)$	Well-adjusted	Maladjusted					
	(33)	(31)		(25)	(31)	(38)	(43)	(42)	(36)	(409)
Movies.										
Mean attendance score	1.3	1.8	3.08	0.6	1.5	1.5	1.4	2.0[2]	1.4	1.5
Percentage of pupils attending frequently	3.2	19.4	2.07	4.2	20.7	11.4	14.6	31.7	5.9	12.8
Percentage of pupils attending infrequently	71.0	35.5	2.99	75.0	37.9	62.9	58.5	31.7	64.7	55.1
Percentage of pupils uninterested	6.1	3.3	0.51	4.0	3.2	10.5	4.7	4.8	8.3	8.6
Organized activities.										
Mean number of activities	1.3	1.4	0.13	1.4	1.1	2.0	1.2	0.9	1.4	1.3
Percentage of extensive participants	18.2	20.0	0.18	8.0	9.7	26.3	14.0	4.8	16.7	12.3
Percentage of non-participants	27.3	26.7	....	20.0	29.0	7.9	34.9	38.1	27.8	24.8
Mean number of hours weekly	5.2	3.8	1.15	3.4	2.7	5.3	3.7	2.8	4.1	4.0
Percentage of eight hours or more	24.2	13.3	1.12	16.0	6.5	26.3	14.0	5.8	19.4	13.0
Percentage of one hour or less	45.5	56.7	0.89	44.0	48.4	26.3	53.5	66.7	50.0	48.4
Percentage with club membership	60.6	70.0	0.78	60.0	64.5	84.2	60.5	78.6	69.4	71.6
Percentage having music lessons	78.8	53.3	2.20	72.0	51.6	79.0	44.2	66.7	52.8	56.5
Percentage remuneratively employed	6.1	0.0	1.45	4.0	12.9	10.5	4.7	16.7	11.1	8.8

[1] Equated for home conditions.
[2] Critical ratio when compared with population 3.4 standard error units.

TABLE II.—SHOWING DIFFERENCES BETWEEN WELL-ADJUSTED AND POORLY ADJUSTED PUPILS WITH RESPECT TO PLAY BEHAVIOR

	Behavior groups										Observed population
	General adjustment			Special traits		Best liked	Fearful	Quick-tempered	Withdrawing		
	Well-adjusted[1] (33)	Maladjusted[1] (31)	Critical ratio $\left(\frac{\text{Diff.}}{\text{SD}_{\text{diff.}}}\right)$	Well-adjusted (25)	Maladjusted (31)	(38)	(43)	(42)	(36)		(409)
Play unit.											
Mean number of school playmates	8.2	6.1	2.65	8.8	6.9	7.7	6.5	7.0	7.5		7.4
Percentage with twelve playmates	12.1	3.3	1.34	12.0	6.5	13.2	2.3	2.4	8.3		7.4
Percentage with three or fewer playmates	3.0	20.0	2.15	0.0	12.9	5.3	11.6	2.4	11.1		6.7
Percentage with large home play group	26.7	16.0	0.97	20.0	11.5	20.0	16.7	18.2	6.5		21.2
Percentage with small home play group	13.3	12.0	0.14	20.0	23.1	20.0	8.3	21.2	22.6		15.7
Percentage preferring large play group	63.6	53.3	0.83	80.0	64.5	68.4	53.5	54.8	47.2		59.9
Mean friendliness score	11.0	11.1	....	10.9	10.8	12.6	11.3	10.2	10.3		10.9
Percentage having many friends	12.1	3.2	1.36	8.0	3.2	10.5	4.7	4.8	0.0		6.7
Percentage having few friends	3.0	3.2	....	0.0	0.0	0.0	0.0	2.4	0.0		2.0
Play behavior.											
Play activity index	8.5	9.0	2.11	8.3	9.7	9.1	8.6	8.8	8.6		8.9
Percentage preferring vigorous plays	0.0	19.4	2.72	8.0	29.0	26.3	14.0	11.9	8.3		16.4
Percentage preferring quiet plays	24.2	22.6	....	36.0	3.2	15.8	32.6	21.4	25.0		23.0

[1] Equated for home conditions.

to the reading data also indicated that a relatively large proportion of these cases expressed a lack of interest in leisure-time reading. Thus, it would appear that this group did not seek the less social types of recreation any more frequently than did the well-adjusted pupils.

Organized Recreational and Educational Activities.—Organized forms of interests in the order of frequency of participation included: Clubs, music lessons and practice, religious classes during the week, scout work, campfire groups, choir practice, dancing school, dramatics, and special classes. The principal results presented here are limited to material collected in the late spring when the pupils were in the high fifth or low sixth grade. On the average, each child devoted four hours weekly to the pursuit of some one or more of these activities. The girls gave more time to these interests and took part in more of them than the boys.

Relatively few differences were found between the well-adjusted and poorly adjusted pupils. From Table II it will be seen that, while an equal proportion among the well-adjusted and poorly adjusted, from a similar socio-economic stratum, participated in several forms of these activities, the well-adjusted tended to give more time to them. The dissimilarities appeared to be a matter of the type of activities in which the groups engaged. The poorly adjusted were more inclined to participate in social forms (such as clubs, scouts, campfire organizations) and in religious activities, while the well-adjusted gave the greatest amount of time to music, dramatics, and dancing classes. But, these differences were never large and the behavior groups were not equally bright.

A contrast of the non-participants with the average pupil in the study and with those who spent at least eight hours weekly in some form of these activities yielded no evidence of an association between the character of classroom behavior and participation in supervised activities. The child who participated extensively had a slightly higher rating on the behavior schedule than the other pupils but the non-participants, likewise, ranked above the average. These differences, however, were not appreciable. The indication is that whether these pupils did or did not have the social and educational advantages which are assumed to accrue from membership in certain types of organizations had little, if any, direct bearing on the patterns of behavior exhibited in the classroom and playground situations during the elementary school years.

Size of Play Unit.—During the period covered by this report, the maximum number of regular playmates on the school playground for the average pupil was around *seven*. The data presented are in terms of the largest number for any semester rather than the average of all records (although both procedures have been employed) in order to prevent discrimination because of transfers near the time the records were taken. The results shown are for the size of the child's play group at school under the most favorable conditions. Information regarding the neighborhood play unit was obtained from the parent at the initiation of the study and again when the child was in the sixth grade. Data from the child was secured at the time of the yearly interviews.

Inspection of these records (see Table II) suggested broader social contacts of the well-adjusted pupils. On the average, the well-adjusted preferred participation with a group and tended to ally themselves with a relatively large number of playmates in both the school and home environments. When behavior was judged in terms of the child's reactions to classroom and playground situations, few of those whose behavior was rated as very satisfactory were reported as playing at school with less than *five* playmates while twenty per cent of the poorly adjusted had *three* playmates as the largest number regularly played with during the four-year period. A similar trend was observed in the records for the personal problem cases. These figures are all exclusive of siblings, but there was no indication that the inclusion of siblings would have changed the relationship.

The school records point to somewhat restricted associations on the playground in the case of poorly adjusted pupils. The criticism against accepting these findings at their face value lies in the possibility of a spurious relationship. It will be recalled that one method of judging behavior was based on the teachers' observations of the child's reactions to the school situations. The results concerning the number of playmates with whom the child associated on the playground may have revealed nothing more than evidences of poor pupil-pupil relationships. That is to say, the maladjusted pupils had difficulties in getting along with their schoolfellows both in the classroom and on the playground. These difficulties may, in turn, have caused either voluntary or group enforced restrictions on the size of their play units. It is very probable that in this complex situation were found many of the overt signs upon which the teachers

relied in making their judgments, and, in this case, large differences would be the expected result.

The fact that similar tendencies were observed in other environments, unprejudiced by school behavior, would indicate that something independent of the child's reactions to school conditions was revealed in these play unit data. Corroborative results were found in the reports from parents and from the pupils. The evidence rests not alone upon the general agreement of all sources of information but also on the definiteness of the trend within a given group. The most frequently occurring descriptions for the well-adjusted pupils concerned their preference for many playmates and their participation in large neighborhood play units. The critical ratios for these proportions indicated reliable differences. There was no evidence, however, of a special interest in restricted play units in the case of the poorly adjusted.

A very interesting point came to light in the study of representatives of withdrawal traits. Classifications of behavior in this instance were based on signs of excessive timidity, sensitiveness, worry, and inferiority feelings as reported by the parent and child. All data collected from the parents and children suggested less extensive participation in group activities by these pupils and less interest in the association with a group (four or more). But, according to the school authorities, their playmate contacts were as broad as those of the average pupil. They did not, then, appear to avoid social contacts, nor did they tend to select nor to prefer the solitary forms of recreation.

The inference from these findings seems to be that the behavior problem groups were less likely to participate in plays and games requiring a large number of players. The well-adjusted, on the other hand, preferred and were usually observed to engage in activities involving a relatively large play group.

Types of Play.—An inventory of play interests included information collected at the time of the yearly individual interviews; answers to a questionnaire relative to play interests which was given during the winter term when the child was in the high fourth or low fifth grade; and, a check list as to frequency of participation and special likes and dislikes, submitted during the late spring at the completion of the fifth grade. From this composite record, it was found that the average pupil named thirty-one different play activities in which he liked to participate and thirteen which he disliked. The character of the

preferred plays was investigated by classifying each game or sport named on a five-step scale, according to the amount of physical energy required. A play activity index was devised by weighting these games by the degree of interest expressed by the child. High scores indicated a preference for vigorously active types of play, and low scores a liking for the quieter forms.

When school behavior was the basis for judging adjustment, the poorly adjusted pupils had the higher activity scores, the differences being large enough to be reasonably certain of the trend. The behavior deviates appeared to be most unlike with respect to their interest in strenuous forms of play, a relatively large number among the unsatisfactory behavior representatives reporting a decided preference for the vigorously active games and sports. The significant point to these results is that the poorly adjusted, on the whole, tended to be less proficient in play performance than the average pupil, according to physical achievement tests of running, throwing, jumping, and balancing. It was not, then, because they could excel in the strenuous activities that they selected them, neither was it a question of their physical robustness—for as a matter of fact they were a less healthy group than the well-adjusted—nor were age nor sex differences involved. The explanation seemed to be a release of pent-up energies and suppressed feelings accumulated in the classroom situation.

Failure was a much more common experience among the maladjusted than it was among the other pupils. Of those who were from a similar socio-economic stratum, there were forty-two per cent of the maladjusted who had failed to pass a grade at least twice during their six or seven years of schooling, with twenty-three per cent of them failing every other year, as opposed to three per cent of the well-adjusted who had failed more than once. Likewise, the maladjusted were much less interested in the classroom projects, fifty-two per cent of them as against three per cent of the well-adjusted reporting a dislike for school activities. The maladjusted were, then, less bright, less interested, and had more frequently faced humiliating experiences of failure and disapproval. It is very likely that the interaction of these several conditions precipitated the more or less explosive activity in the spontaneous interests of the playground.

That this was at least part of the explanation was indicated by the results from the special problem cases and from findings with respect to pupils who preferred the strenuous forms of play as contrasted with those who liked the less vigorous games. In the former

instance, without exception the poorly adjusted, by criteria other than school behavior, had a lower average play activity score than the general run of the pupils surveyed, with fewer of them expressing a preference for very active forms of play. None of these differences, however, were large enough to eliminate chance variations.

The play activity extremes included the ten per cent of the boys and of the girls who expressed a decided preference for the very active plays and games and the same proportion who selected the less active plays. An inspection of the comparative results indicated that these pupils were most unlike each other in their ability to get along well in school. The boys and the girls who preferred the vigorous forms of play were less bright, less successful in school work, and had lower ratings in behavior than either the average pupil under observation or those preferring the less active games and plays. It is not unlikely that these results throw into relief an important technique utilized by the child in compensating for poor school adjustments. Unhampered on the playground by the handicaps of the classroom, the child puts all of his energies into an effort to remove the humiliations of the day's experiences. While it is possible that a certain amount of rowdiness in play is usually associated with unsatisfactory behavior adjustments, it is highly probable that this aspect of play may be merely an outgrowth of the lack of adaptation of school procedures to meet the needs of individual pupils.

GENERAL RELATIONSHIPS AND CONCLUSIONS

The results, on the whole, appear to show that when children are from a similar socio-economic level, there is little evidence of a general association between the activities to which they give their "free hours" and the adequacy or inadequacy of their overt behavior adjustments. When a number of measures of behavior were correlated with measures of recreation, the coefficients were all under .35, seventy per cent of them being less than .20. In the first column of Table III, are shown a few coefficients of mean square contingency (C) and in the remaining columns the product-moment coefficients (r). The last two columns present some results when either age or IQ was held constant. There was an indication of a low positive relation between desirable behavior traits and participation in supervised forms of educational and recreational activities and of a slight but negative relation with both a preference for strenuous games and plays and frequency of attendance at the motion picture show. In the case of the other measures the

TABLE III.—CORRELATIONS BETWEEN RECREATIONAL ACTIVITIES AND MEASURES OF BEHAVIOR

Measures of recreation	Measures of behavior	Correlation coefficients			
		C	r	Partials	
				Age	IQ
Play proficiency or athletic ability	General adjustment coefficient	.....	.150 ± .032	.083 ± .033	
	Social traits	.....	.163 ± .032	.109 ± .033	
	Emotional traits	.....	−.084 ± .033		
	Social recognition	.....	.253 ± .031	.210 ± .032	
Play activity	General adjustment coefficient	.....	−.298 ± .030	−.280 ± .030	
	Conduct mark	.....	−.290 ± .030		
	Social traits	.....	−.101 ± .033		
	Emotional traits	.....	−.136 ± .033		
	Social recognition	.....	.002 ± .033		
Play versatility	General adjustment coefficient	.....	−.026 ± .033		
Play unit (school)	General adjustment coefficient	.....	.182 ± .033		
	Emotional traits	.....	.038 ± .033		
	Social recognition	.....	.066 ± .033		
Supervised activities:					
Number	General adjustment coefficient	.....	.161 ± .032	.....	.115 ± .033
Time (weekly)	General adjustment coefficient	.231	.154 ± .033	.....	.108 ± .033
	Social traits	.278	.151 ± .032	.....	.105 ± .033
	Social recognition	.310			
Leisure-time reading (amount)	General adjustment coefficient	.....	.175 ± .032	.....	.070 ± .033
	Social traits	.....	.095 ± .033		
	Emotional traits	.....	.047 ± .033		
Movie attendance	General adjustment coefficient	−.241			
	Emotional traits	−.311			
	Social recognition	−.263			

coefficients either varied in direction with the different behavior categories or approximated zero in size.

The implication from the findings of this investigation seems to be that what children of elementary school age do with their after-school hours is not an important conditioning factor in their personal adjustments. It is probable that the value of recreations for wholesome personality may be found in the nature of the individual child's reaction to the activity and in the interrelation of various intrinsic and extrinsic developmental conditions rather than in the presence or absence of certain contacts.

WHO OFFERS WHAT IN THE PSYCHOLOGY OF EDUCATION?

VEST C. MYERS

Southeast Missouri Teachers College, Cape Girardeau, Missouri

Psychology is today a word to conjure with. The minister refers to the psychology of religion; the sales manager talks of the psychology of advertising; the scholarly sociologist alludes to national or racial psychology; the high school student speaks of the bad psychology of a certain football play, while the man on the street, not to be left behind, makes use of the terms good psychology or poor psychology as he attempts to interpret some problem dealing with human relationships.

One sometimes wonders if psychology having won the world will lose its own soul in that it may become a component part of other fields of knowledge and lose its identity as a subject of the curriculum. As teachers we may be interested in knowing just what phases of educational psychology are emphasized by institutions or departments of institutions, which are especially interested in the training of teachers. What are the present day tendencies as reflected by the catalogues of representative institutions?

The writer recently made a survey of the offerings of a large group of universities, liberal arts, colleges, and teachers colleges, in the field of educational psychology and closely related subjects. The material used in the study consisted of the catalogues of fifty universities, thirty of which were state universities; of twenty-six liberal arts colleges, located in nineteen states, and of one hundred teachers colleges, representing thirty-nine states. It is believed that this group of one hundred seventy-six institutions, widely distributed over the whole country, should give a fairly accurate idea of the general nature of the courses offered in the psychology of education.

The table at the top of page 469 lists the courses offered by the three types of schools and the percentage of schools offering each course.

In the study of university catalogues only the courses in psychology associated with the educational departments of the institutions were included. Certain courses catalogued as Systems of Psychology, Comparative Psychology, Experimental Psychology, History of Psychology, and Physiological Psychology were not considered since they have little direct relationship with the field of Education. In

Courses offered	Universities		Liberal arts colleges		Teachers colleges	
	Number	Percentages	Number	Percentages	Number	Percentages
Educational psychology	50	100	23	88	100	100
General psychology	50	100	20	77	31	31
Child psychology	40	80	14	54	68	68
Abnormal psychology	43	86	11	42	29	29
Adolescent psychology	16	32	6	23	18	18
Social psychology	31	62	11	42	33	33
Psychology of elementary subjects	21	42	2	8	20	20
Psychology of high school subjects	17	34	4	15	17	17
Guidance	37	74	6	23	41	41
Mental hygiene	24	48	8	31	23	23
Character education	23	46	2	8	18	18
Tests and measurements	47	94	16	62	97	97
Educational statistics	46	92	6	23	26	26

computing percentages, fractions of less than one-half were omitted; otherwise, they were counted as one.

In twenty-eight of the universities and twenty-one of the teachers colleges, general psychology is a prerequisite to educational psychology. Generally the course in general psychology or educational psychology is a pre-requisite for the other courses listed in the table. Several of the schools studied probably covered a considerable part of the field of Social Psychology in courses listed as Sociology. A number of the liberal arts college, judging from the catalogue description of the courses, offered considerable work in character education in courses listed as Personality, Human Relationships, etc.

Eleven of the eighteen teachers colleges offering work in character education are west of the state of Missouri.

The generalizations from such a study as the present one must of necessity be somewhat limited. However, certain trends seem to be apparent. The psychology of education, in its various phases, is a subject of outstanding importance in all of the three types of schools. General Psychology appears to be considered of greater importance by the educational departments of the universities and liberal arts

colleges than by the teachers colleges. Much greater interest is being placed on child psychology than on adolescent psychology. Guidance, though a relatively new subject, has been accepted as one of the important fields of study. The mental health program is receiving serious attention as is shown by the interest in mental hygiene and abnormal psychology. The subject of tests and measurements has been accepted by a large majority of the schools. Considerable attention is being given to the study of character education.

THE EFFECT OF A GROUP OF SOCIAL STIMULI UPON ATTITUDES[1]

IVA COX GARDNER

Baylor University, Waco, Texas

INTRODUCTION

In this series of experimental studies, an attempt has been made to find the effect of a group of social stimuli upon attitudes. One thousand two hundred thirty-six junior high school students and college freshmen in Waco, Texas, served as subjects for these experiments. The Peterson-Thurstone scale on *Attitude Toward War* and the Smith-Thurstone scale on *Attitude Toward Prohibition*, Forms *A* and *B* were used for test and retest. The social stimuli consisted of lectures, stories and chalk talks. All stimuli in a series were presented on one side of the issue. Briefly, the procedure has been to measure the attitudes of a group of students by means of an attitude scale, then to present a stimulus which might affect the attitudes on the issue in question, and, on the following day, to measure again the attitudes of the students; a week later to present another stimulus on the same side of the issue as before, following this the next day with another form of the attitude scale; again a week later to present a third stimulus on the same side of the issue, following it the next day with the fourth presentation of the scale. Since there were only two forms of the scales available, it was necessary to present each of the forms twice.

The experimental groups consisted of two eighth grade groups in junior high school (the two schools being located in different sections of the city), one seventh grade group, and one college freshman group. The control group was composed of ninth grade high school students. The schools from which all these groups were selected are located in Waco, Texas—a city of approximately sixty thousand population.

PROCEDURE AND RESULTS

The experiments on the control group will be reported first, since the two forms of the scale on *Attitude Toward War* were found not

[1] The essential portion of a dissertation presented to the Department of Psychology of the University of Chicago in candidacy for the degree of Doctor of Philosophy.

comparable. This condition made necessary certain corrections throughout the study.

At the beginning of each experiment, the students were informed as to the number of times they would be asked to mark the scale. This was done in order to prevent dissatisfaction when the scale was presented several times.

Each Thursday, beginning the first week in April, and running for four consecutive weeks, the attitude scale, *Attitude Toward War* was presented to over two hundred ninth grade students in the Waco Senior High School. However, only one hundred eighty students remained throughout the entire experiment, and only these were included in this study. The first week, Form *A* of the scale was given, the second week Form *B*, the third week Form *A*, and the fourth week Form *B*. No social stimuli were given to this group.

The statistical facts about the frequency distributions of the control group after the corrections referred to above are as follows:

TABLE I.—SHOWING MEANS, STANDARD DEVIATIONS, PROBABLE ERRORS OF MEANS, AND CORRELATIONS, TOGETHER WITH DIFFERENCES BETWEEN MEANS, AND THE RATIO OF THE DIFFERENCES TO THEIR PROBABLE ERRORS
Waco Senior High School—Control Group

$M_1 = 4.653$	$D_{M_1-M_2} = -.085$	$PE_{M_1} = .059$	$r_{12} = .88$
$M_2 = 4.738$	$D_{M_2-M_3} = .005$	$PE_{M_2} = .055$	$r_{34} = .86$
$M_3 = 4.733$	$D_{M_3-M_4} = .050$	$PE_{M_3} = .064$	$r_{23} = .87$
$M_4 = 4.683$	$D_{M_1-M_4} = -.030$	$PE_{M_4} = .061$	
$\sigma_1 = 1.19$	$PE_{D_{12}} = .028$	$D/PE_{D_{12}} = 3.04$	$n = 180$
$\sigma_2 = 1.13$	$PE_{D_{23}} = .032$	$D/PE_{D_{23}} = .16$	
$\sigma_3 = 1.29$	$PE_{D_{34}} = .033$	$D/PE_{D_{34}} = 1.52$	
$\sigma_4 = 1.23$	$PE_{D_{14}} = .031$	$D/PE_{D_{14}} = .97$	

From the above table, it is seen that the ratio of the difference to the probable error of the difference in each instance is too small to be statistically reliable, showing that there was no significant change in the attitudes of the group over the period of testing.

In the experimental group, two of the experiments were conducted simultaneously in February, 1933, with eighth grade students. One of the groups was at the South Junior High School, the other at the West Junior High School. A definite schedule of procedure was followed in each school, allowing a lapse of a week between the successive presentations of both the scales and the stimuli.

The first stimulus was presented by means of a lecture. The lecturer[1] used appeals both to reason and emotion. In substance the same lecture was given at the two schools, except for the inclusion at South Junior of two emotional stories which tended to glorify war.

The second stimulus, a pacifistic story by another speaker,[2] contained elements of reason, with much that was emotional in its appeal for peace.

The last stimulus at South Junior was a chalk talk by a third speaker.[3] This speaker brought out facts about the cost of war in money and lives, and also stressed the profit-making motive.

Just before the presentation of this chalk talk, the play *The Valiant* was given in the school auditorium—a play strongly emphasizing the glory of dying for one's country. The scale given the following day showed some change in attitude toward peace, but the movement on the scale was very small.

On the day for giving the third stimulus to the West Junior group, this speaker was not available, and the experimenter was forced to present the stimulus. This was done by showing in tableau form how little is gained when two boys fight over an object claimed by both. Emphasis was placed on the cost of war, some paying with their lives, others with their money, for years after the war is over.

TABLE II.—SHOWING MEANS, STANDARD DEVIATIONS, PROBABLE ERRORS OF MEANS, AND CORRELATIONS, TOGETHER WITH DIFFERENCES BETWEEN MEANS, AND THE RATIO OF THE DIFFERENCES TO THEIR PROBABLE ERRORS

South Junior High School—Eighth Grade

M_1 before $S = 4.742$	$D_{M_1-M_2} = .065$	$PE_{M_1} = .049$	$r_{12} = .838$
M_2 after $S_1 = 4.677$	$D_{M_2-M_3} = .275$	$PE_{M_2} = .048$	$r_{23} = .835$
M_3 after $S_2 = 4.402$	$D_{M_3-M_4} = .168$	$PE_{M_3} = .053$	$r_{34} = .90$
M_4 after $S_3 = 4.234$	$D_{M_1-M_4} = .508$	$PE_{M_4} = .051$	$r_{14} = .88$

$\sigma_1 = 1.15$	$PE_{D_{12}} = .028$	$D/PE_{D_{12}} = 2.32$	$n = 247$
$\sigma_2 = 1.13$	$PE_{D_{23}} = .029$	$D/PE_{D_{23}} = 9.48$	
$\sigma_3 = 1.25$	$PE_{D_{34}} = .023$	$D/PE_{D_{34}} = 7.30$	
$\sigma_4 = 1.21$	$PE_{D_{14}} = .025$	$D/PE_{D_{14}} = 20.32$	

In all the frequency tables in this study on *Attitude Toward War*, the scores of Form *B* have been made equivalent to Form *A*. Also,

[1] Mrs. J. M. Dawson.
[2] Miss Lois Shaw.
[3] Mrs. George Miller.

a low score indicates an attitude favorable to peace. Statistical facts about the frequency distributions of these two groups are shown in Tables II and III.

TABLE III.—SHOWING MEANS, STANDARD DEVIATIONS, PROBABLE ERRORS OF MEANS, AND CORRELATIONS, TOGETHER WITH DIFFERENCES BETWEEN MEANS, AND THE RATIO OF THE DIFFERENCES TO THEIR PROBABLE ERRORS
West Junior High School—Eighth Grade

M_1 before $S = 4.698$	$D_{M_1-M_2} = .757$	$PE_{M_1} = .047$	$r_{12} = .85$
M_2 after $S_1 = 3.941$	$D_{M_2-M_3} = .091$	$PE_{M_2} = .041$	$r_{23} = .89$
M_3 after $S_2 = 3.850$	$D_{M_3-M_4} = .865$	$PE_{M_3} = .049$	$r_{34} = .80$
M_4 after $S_3 = 2.985$	$D_{M_1-M_4} = 1.713$	$PE_{M_4} = .017$	$r_{14} = .67$

$\sigma_1 = 1.19$	$PE_{D_{12}} = .024$	$D/PE_{D_{12}} = 31.54$	$n = 291$
$\sigma_2 = 1.05$	$PE_{D_{23}} = .022$	$D/PE_{D_{23}} = 4.14$	
$\sigma_3 = 1.25$	$PE_{D_{34}} = .035$	$D/PE_{D_{34}} = 24.71$	
$\sigma_4 = .45$	$PE_{D_{14}} = .037$	$D/PE_{D_{14}} = 46.30$	

In the South Junior group, as seen from the above table, the conclusion that the three social stimuli made the students more favorable toward peace seems justified. The effect from all stimuli is cumulative, and one strong stimulus does not have the effect of several successive stimuli, though counterbalancing stimuli have their effect.

In the West Junior group, the same cumulative effect of the three stimuli is seen, with a much more decided change in attitude than was found in the South Junior group and a more consistent lowering of the mean score with the presentation of the successive stimuli.

This difference in the effect on the two groups is noted particularly following the first and third stimuli. As noted above, the first lecture to the two groups was the same, except the inclusion in the address to the South Junior group of two emotional stories which tended to glorify war. These stories constituted a conflicting element in the stimulus that largely offset the speaker's arguments for peace, and the movement on the scale as a result of this stimulus was too small to be statistically reliable. The play presented immediately preceding the third stimulus to this same group, as noted above, also largely offset the third speaker's arguments against war, and the resulting movement on the scale was very small. The second stimulus, which had no such counteracting elements, showed a more decided change toward peace. On the whole, the South Junior group did not show a decided change in attitude, but the conflicting stimuli no doubt kept

this group within one scale step of the first mean score. This group, we might say, served as a sort of control, the conflicting stimuli preventing any very great change in attitude.

In the West Junior group, where the conditions were more perfectly controlled, there was a decided movement on the scale toward peace with the presentation of each stimulus. In this group the movement on the scale may be interpreted as a change in attitude toward war from neutrality or near neutrality to strong opposition to war, the mean score being used for this interpretation. This group also became more homogeneous after the presentation of each stimulus, as indicated by the standard deviations.

An attitude scale which will show changes in the direction of a given stimulus, and then again will show little or no change when conflicting stimuli are given, may be said to be a valid scale for measuring attitudes. The control group at the Senior High School leads to the same conclusion, as there was no significant change in attitudes where no social stimuli were presented.

TABLE IV.—SHOWING MEANS, STANDARD DEVIATIONS, PROBABLE ERRORS OF MEANS, AND CORRELATIONS, TOGETHER WITH DIFFERENCES BETWEEN MEANS, AND THE RATIO OF THE DIFFERENCES TO THEIR PROBABLE ERRORS
West Junior High School—Seventh Grade

M_1 before $S = 4.724$	$D_{M_1-M_2} = .976$	$PE_{M_1} = .040$	$r_{12} = .82$
M_2 after $S_1 = 3.748$	$D_{M_2-M_3} = .561$	$PE_{M_2} = .031$	$r_{23} = .72$
M_3 after $S_2 = 3.187$	$D_{M_3-M_4} = .253$	$PE_{M_3} = .020$	$r_{34} = .61$
M_4 after $S_3 = 2.934$	$D_{M_1-M_4} = 1.790$	$PE_{M_4} = .022$	$r_{14} = .71$

$\sigma_1 = 1.02$	$PE_{D_{12}} = .022$	$D/PE_{D_{12}} = 44.36$	
$\sigma_2 = .81$	$PE_{D_{23}} = .021$	$D/PE_{D_{23}} = 26.71$	$n = 290$
$\sigma_3 = .53$	$PE_{D_{34}} = .018$	$D/PE_{D_{34}} = 14.05$	
$\sigma_4 = .57$	$PE_{D_{14}} = .028$	$D/PE_{D_{14}} = 63.92$	

Following these two experiments with eighth grade groups, another experiment was undertaken with the seventh grade group of West Junior High School, beginning the first week in April, 1933. The same schedule and procedure were followed as in the former experiments, using the same scale on *Attitude Toward War*. Two stimuli used in the former experiment with the West Junior eighth grade were used in this one—the story and the experimenter's talk—the former verbatim, the latter practically so. It was necessary to secure a

different speaker[1] to present the first stimulus, as the former speaker was not available at this time.

This speaker's manner of presentation was one of the factors in the strength of his lecture. He called attention to the causes of war, stressing greed and covetousness. He pointed out the utter uselessness of war as seen from its results, and clearly presented the possibility of settlement of all difficulties by arbitration.

The statistical facts about the frequency distributions of this group are shown in Table IV.

The above table shows that these students were moved decidedly in their attitudes toward peace by the presentation of the three social stimuli. The first stimulus shows the greatest movement on the scale, but the other two reveal an added movement toward pacifism. The movement on the scale in this group may be interpreted like that of the West Junior eighth grade—as a change in attitude toward war from neutrality or near neutrality to strong opposition. This group also shows growing homogeneity after the presentation of the successive stimuli. The ages of these junior high school students may have made them more susceptible to the stimuli, though this is not conclusive from these experiments.

During the first week in April, 1933, an experiment was begun with the freshman class of Baylor University. A group of about two hundred and fifty began, but only two hundred twenty-eight remained throughout the experiment. The Smith-Thurstone scale on *Attitude Toward Prohibition*, Forms *A* and *B* were used in this experiment. The stimuli presented were favorable to prohibition.

On Wednesday of the first week in April, Form *A* of this scale was given to the group. The following Tuesday, the president of the university[2] gave a lecture on prohibition to the entire student body in chapel, making a strong appeal in favor of prohibition. On the following day, when only the freshmen attended chapel, Form *B* of the prohibition scale was given.

The second stimulus was given during the following week by a group of students under the direction of one of the student leaders.[3] In the group was a law student,[4] a pre-medical student,[5] and a min-

[1] Mr. Caso March.
[2] Hon. Pat M. Neff.
[3] Mr. William Stone.
[4] Mr. Caso March.
[5] Mr. William Bird.

isterial student.[1] The students were known prohibitionists on the campus and were respected for their views, though many did not agree with them. Each made an appeal from his particular standpoint. This stimulus was followed by the second presentation of Form *A* of the scale.

At this time the prohibition question was a political issue. Following the series of speeches by members of the student body, the anti-prohibitionists on the campus became active, and the experimenter decided it best not to present a third stimulus.

On this scale, a high score represents an attitude favorable to prohibition. The statistical facts about the frequency distributions of this group are shown in the following table:

TABLE V.—SHOWING MEANS, STANDARD DEVIATIONS, PROBABLE ERRORS OF MEANS, AND CORRELATIONS, TOGETHER WITH DIFFERENCES BETWEEN MEANS, AND THE RATIO OF THE DIFFERENCES TO THEIR PROBABLE ERRORS

Baylor Freshmen

M_1 before $S = 5.798$ M_2 after $S_1 = 6.595$ M_3 after $S_2 = 7.147$	$D_{M_2-M_1} = .797$ $D_{M_3-M_2} = .552$ $D_{M_3-M_1} = 1.349$	$PE_{M_1} = .079$ $PE_{M_2} = .072$ $PE_{M_3} = .068$	$r_{12} = .90$ $r_{23} = .91$ $r_{13} = .90$
$\sigma_1 = 1.78$ $\sigma_2 = 1.62$ $\sigma_3 = 1.54$	$PE_{D_{12}} = .034$ $PE_{D_{23}} = .029$ $PE_{D_{13}} = .034$	$D/PE_{D_{12}} = 23.44$ $D/PE_{D_{23}} = 19.04$ $D/PE_{D_{13}} = 39.68$	$n = 228$

As seen from the table, the attitudes of the students considered as a group were changed significantly toward prohibition. The effect of the three stimuli is cumulative, as was true in the other experiments, and the same trend is shown as in the other groups. At the time of this experiment, prohibition had become somewhat unpopular among the masses. However, the scores revealed that the stimuli were effective in modifying the attitudes of the students. The mean score in the beginning indicated an attitude of indifference toward prohibition. After the two stimuli, they were at least mildly favorable, almost reaching the scale value indicating "strongly favorable" to prohibition.

This group did not become as homogeneous as the former groups after being presented with the stimuli on peace. This fact can be observed from the standard deviations and the range. The standard deviation changes but little.

[1] Mr. Leon Gorsline.

CONCLUSIONS

From the experiments reported in this paper, the cumulative effect of a group of stimuli upon an attitude has been demonstrated. The general plan of the experiment was to measure the attitudes of a group of subjects by means of an attitude scale, then to present a group of social stimuli which had been judged as having affective value on the issue in question, the attitudes of the subjects being measured before and after the presentation of each stimulus. The experimenter arbitrarily limited to three the number of stimuli to be used. The social stimuli were of a lecture or story type. An interval of one week was allowed between the presentation of each scale and the succeeding one; also an interval of a week elapsed between the presentation of the successive stimuli.

The issues studied were those of war and prohibition. The scale on *Attitude Toward Prohibition* was used only in Baylor University; because of the political issues involved, it was not deemed wise to use it in the public schools. Instead, the scale on *Attitude Toward War* was used there. On this scale, the stimuli which were presented were pacifistic in nature. Two of the groups showed a decided change toward pacifism after the presentation of the three stimuli.

The college group did not show so much change in attitude as did the high school groups. This might be due to the fact that they were older and more sophisticated, and that their attitudes were set and not so easily changed. Or it might be that the issue under consideration in the college group is one on which more firmly fixed convictions are held.

The findings in this experiment are in keeping with Peterson and Thurstone's findings in regard to the motion pictures: *viz.*, that a combination of two pictures on the same issue, neither of which has a measurable effect, may show a significant change in attitude toward the issue in question.

PUBLICATIONS RECEIVED

[Mention in this section does not preclude a work from receiving extended notice in a later issue.]

GORDON W. ALLPORT and PHILIP E. VERNON. *Studies in Expressive Movement.* New York: The Macmillan Co., 1933, pp. 269.

MILTON ANDERSON. *The Modern Goliath.* A Study of Talking Pictures. Los Angeles: David Press, 1935, pp. 90.

RALPH C. BEDELL. *The Relationship between the Ability to Recall and the Ability to Infer in Specific Learning Situations.* Kirksville, Mo.: Northeast Missouri State Teachers College, 1934, pp. 56. (paper.)

C. S. BLUEMEL. *Stammering and Allied Disorders.* New York: The Macmillan Co., 1935, pp. 182.

LEROY E. BOWMAN. *Community Programs for Summer Play Schools.* New York: Child Study Association of America, 1935, pp. 48. (paper.)

CYRIL BURT. *The Subnormal Mind.* New York: Oxford University Press, 1935, pp. 368.

MORSE ADAMS CARTWRIGHT. *Ten Years of Adult Education.* New York: The Macmillan Co., 1935, pp. 220.

LUELLA COLE and JESSIE MARY FERGUSON. *Students' Guide to Efficient Study.* New York: Farrar and Rinehart, Inc., 1935, pp. 38. (paper.)

Conference on Three Special Problems in Guidance. Professional Standards in Guidance. Planning Occupational Futures. The Economic Problem Group. New York: Department of Psychology, Fordham University Graduate School, 1935, pp. 136. (paper.)

EDMUND S. CONKLIN. *Principles of Adolescent Psychology.* New York: Henry Holt & Co., 1935, pp. 437.

ETHEL L. CORNELL and WARREN W. COXE. *A Performance Ability Scale.* Examination Manual. Yonkers, N. Y.: World Book Company, 1934, pp. 88. *Cornell-Coxe Performance Ability Scale.* Individual Record Blank. Yonkers, N. Y.: World Book Co., 1934, pp. 4. (paper.)

EDGAR DALE. *The Content of Motion Pictures.* New York: The Macmillan Co., 1935, pp. 281.

ROBERT A. DAVIS. *Psychology of Learning.* New York: McGraw-Hill Book Co., Inc., 1935, pp. 489.

JOHN MORRIS DORSEY. *The Foundations of Human Nature.* New York: Longmans, Green and Co., 1935, pp. 488.

F. M. EARLE and J. KILGOUR. *A Vocational Guidance Research in Fife.* London: National Institute of Industrial Psychology, 1935, pp. 101. (paper.)

FAYETTE H. ELWELL and JOHN GUY FOWLKES. *Instructional Tests in Bookkeeping for High Schools and Colleges.* Yonkers, N. Y.: World Book Co., 1934, pp. 92. (paper.)

MARIAN J. FITZ-SIMONS. *Some Parent-child Relationships.* New York: Bureau of Publications, Teachers College, Columbia University, 1935, pp. 162.

JOHN CLEMANS FLANAGAN. *Factor Analysis in the Study of Personality.*

Stanford University, Calif.: Stanford University Press, 1935, pp. 103. (paper.)

ARTHUR I. GATES. *The Improvement of Reading.* Revised Edition. New York: The Macmillan Co., 1935, pp. 668.

COLEMAN R. GRIFFITH. *An Introduction to Educational Psychology.* New York: Farrar & Rinehart, Inc., 1935, pp. 754.

JOSEPH HARRY GRIFFITHS. *The Psychology of Human Behavior.* New York: Farrar and Rinehart, Inc., 1935, pp. 515.

E. R. GUTHRIE. *The Psychology of Learning.* New York: Harper & Brothers, 1935, 258.

HORNELL HART and ELLA B. HART. *Personality and the Family.* Boston: D. C. Heath and Co., 1935, pp. 381.

CLAUDE EDWIN HEATON. *Modern Motherhood.* New York: Farrar and Rinehart, Inc., 1935, pp. 271.

KATE HEVNER. *Visual Outline of Psychology.* New York: Longmans, Green & Co., 1935, pp. 97. (paper.)

Historical Approach to Methods of Teaching the Social Studies, 1935, *the.* Fifth Yearbook, The National Council for the Social Studies. Philadelphia: McKinley Publishing Co., 1935, pp. 204. (paper.)

KARL J. HOLZINGER. *Preliminary Report on Spearman-Holzinger Unitary Trait Study.* Five Pamphlets, pp. 27, 26, 37, 78, and 64. Chicago: Statistical Laboratory, Department of Education, The University of Chicago, 1934 and 1935.

RICHARD AMARAL HOWDEN. *The Man in the Street and the New Psychology.* New York: Oxford University Press, 1935, pp. 77.

KURT KOFFKA. *Principles of Gestalt Psychology.* New York: Harcourt, Brace and Co., 1935, pp. 720.

EDNA E. KRAMER. *A First Course in Educational Statistics.* New York: John Wiley & Sons, Inc., 1935, pp. 212.

H. LAUGIER, HENRI PIERON and OTHERS. *Etudes Docimologiques Sur le Perfectionnement des Emamens et Concours.* Paris: Conservatoire National Des Arts et Metiers, 1934, pp. 89. (paper.)

R. W. LEIGHTON. *Studies of Laboratory Methods of Teaching.* Robert H. Seashore. *Qualitative Aspects in the Improvement of Science Teaching.* Ann Arbor, Mich.: Edward Brothers, Inc., 1935, pp. 184. (paper.)

EDNA BRIDGE LEINING. *Millions of Years in a Winter.* New York: Bureau of Publications, Teachers College, Columbia University, 1935, pp. 197.

HANNAH LOGASA. *Historical Fiction and Other Reading References for History Classes in Junior and Senior High Schools.* Philadelphia: McKinley Pub. Co., 1934, pp. 144. (paper.)

JOHN A. LONG and PETER SANDIFORD. *The Validation of Test Items.* Toronto: Department of Educational Research, University of Toronto, 1935, pp. 126. (paper.)